المعهد الملكي للثقافة الأمازيغية
INSTITUT ROYAL DE LA CULTURE AMAZIGHE

CITISI'2018

Conférence Internationale

TECHNOLOGIE, INNOVATION SYSTEME D'INFORMATION

14 et 15 juillet 2018

ACTES DE LA SESSION THEMATIQUE
TRAITEMENT AUTOMATIQUE DES LANGUES

Coordination

Fadoua ATAA ALLAH

Siham BOULAKNADEL

CONFERENCE INTERNATIONALE

TECHNOLOGIE, INNOVATION SYSTEME D'INFORMATION

14 - 15 juillet 2018

ACTES DE LA SESSION THEMATIQUE

TRAITEMENT AUTOMATIQUE DES LANGUES

Coordination

Fadoua Ataa Allah & Siham Boulaknadel

CONFERENCE INTERNATIONALE TECHNOLOGIE, INNOVATION SYSTEME D'INFORMATION

Centre des Etudes Informatiques, des Systèmes
d'Information et de Communication

CITISI

CONFERENCE INTERNATIONALE TECHNOLOGIE, INNOVATION SYSTEME D'INFORMATION

14 et 15 juillet 2018

ACTES DE LA SESSION THEMATIQUE

TRAITEMENT AUTOMATIQUE DES LANGUES

Coordination
Fadoua ATAA ALLAH & Siham BOULAKNADEL

Publications de l'Institut Royal de la Culture Amazighe
Centre des Etudes Informatiques, des Systèmes d'Information et de Communication

Série : Colloques et séminaires N° 65

Titre
CONFERENCE INTERNATIONALE TECHNOLOGIE, INNOVATION, SYSTEME D'INFORMATION
Actes de la session thématique
TRAITEMENT AUTOMATIQUE DES LANGUES

Coordination
Fadoua ATAA ALLAH & Siham BOULAKNADEL

Editeur
Institut Royal de la Culture Amazighe

Imprimerie
Edition et impression Bouregreg - Rabat

Dépôt légal
2022MO0133

ISBN
978-9920-739-52-8

Copyright
©IRCAM

PREFACE

La sélection d'articles publiés dans ce recueil constitue les actes de la session thématique sur le traitement automatique des langues. Elle a été organisée par l'Institut Royal de la Culture Amazighe, sous l'initiative du Centre des Etudes Informatiques, des Systèmes d'Information et de Communication, en partenariat avec la Faculté des Sciences de Kénitra, l'Ecole Nationale des Sciences Appliquées et le Club des Docteurs Science et Technologie. Cette session thématique sur le traitement automatique des langues s'inscrivait dans le cadre des activités scientifiques de la 2ème édition de la conférence Internationale Technologie, Innovation, Système d'Information (CITISI) qui s'est tenue les 14 et 15 juillet 2018, à la Faculté des Sciences de Kénitra.

La session sur le traitement automatique des langues a réuni des académiciens de plusieurs établissements représentant les universités marocaines et étrangères. Les travaux de cette rencontre scientifique se sont articulés autour de la question essentielle du traitement des langues par le biais des sciences du numérique. Elle visait la sensibilisation de la communauté académique aux enjeux et aux difficultés rencontrées dans le traitement des langues et la mise au point des méthodologies simples et économiques pour l'élaboration de ressources et outils.

Deux thèmes majeurs ont ponctué cette manifestation : le premier a trait à l'apprentissage médiatisé par la technologie et le second à l'élaboration de ressources et outils linguistiques.

Nous remercions vivement les auteurs pour leurs contributions, le comité scientifique pour la qualité de leurs lectures et évaluations ainsi que le comité d'organisation pour son travail efficace. Nous remercions également l'Institut Royal de la Culture Amazighe, la Faculté des Sciences de Kénitra, l'Ecole Nationale des Sciences Appliquées et le Club des Docteurs Science et Technologie pour leurs aides financières et soutiens divers. Sans ces soutiens, comment peut-il réaliser un tel travail ?

Comité scientifique

J. Abouchabaka (FS-Kénitra, Maroc)

M. Ait Kerroum (ENCG-Kénitra, Maroc)

El. B Ameur (FS-Kénitra, Maroc)

M. Amraoui (EST-Salé, Maroc)

F. Ataa Allah (IRCAM, Maroc)

S. Boulaknadel (IRCAM, Maroc)

S. Bourekkadi (SUP'MM Tanger, Maroc)

F. G. Caballero (Univ-Grenade, Espagne)

H. M. Chaoui (ENSA-Kénitra, Maroc)

Y. Chaabi (IRCAM, Maroc)

A. El Moudden (ENCG-Kénitra, Maroc)

Y. Fakhri (FS-Kénitra, Maroc)

R. Jourani (FSJES-Tétouan, Maroc)

S. Khoulji (ENSA-Tétouan, Maroc)

A. Moumen (ENSA-Kénitra, Maroc)

N. Rafalia (FS-Kénitra, Maroc)

D. P. Ruiz (Univ-Grenade, Espagne)

A. Souhar (FS-Kénitra, Maroc)

Comité d'organisation

F. Ataa Allah (IRCAM, Maroc)

S. Boulaknadel (IRCAM, Maroc)

Y. Chaabi (IRCAM, Maroc)

J. Frain (IRCAM, Maroc)

TABLE DES MATIERES

Préface.....V

Detection Learner Style Through Genetic Algorithm.....1
F.Z. HIBBI, O. ABDOUN, E.K. HAIMOUDI

Analyse Sémantique de Conversations pour l'Identification des Profils Comportementaux sur Facebook.....11
Y. CHAABI, K. LEKDIOUI, W. MESTADI

Building Virtual Worlds for Amazigh Game based Learning.....33
Y. TAZOUTI, S. BOULAKNADEL, Y. FAKHRI

Conception d'un Système d'Identification de Stratégie d'Apprentissage à base des Représentations Analogiques.....49
W. MESTADI, Y. CHAABI, K. NAFIL

Pedagogical Approaches to Teaching and Learning Entrepreneurship Education in the Polydisciplinary Faculty of LARACH.....65
S. BOUREKKADI, A. BABOUNIA, K. SLIMANI, M. EL KANDILI, S. KHOULJI

A Dialogue-System Using a Quranic Ontology.....75
F. BENDJAMAA, N. TALEB

A Brief Survey on Named Entity Recognition in Amazigh Language.....89
M. TALHA, S. BOULAKNADEL, A. HAMMOUCH

Text Mining : de l'Information Textuelle à la Connaissance.....99
K. SELLAMY, Y. FAKHRI, K. HAFED

Automatic Recognition of Arabic Handwritten Words.....111
A. LAMSAF, M. AIT KERROUM, S. BOULAKNADEL, C. SAMIR

Detection Learner Style Through Genetic Algorithm

HIBBI Fatima-Zohra, ABDOUN Otman, EL KHATIR Haimoud

Department of computer science, Abdelmalek Essaadi University, Morocco

Fatima.zohra.official@gmail.com
abdoun.otman@gmail.com
helkhatir@gmail.com

Abstract

Users in a class have different interactions in a particular situation in front of the machine. The tutor or face-to-face trainer has the ability to adapt his content according to the immediate feedback he can get from the learners. If one of the students gets bored or stalls, the trainer can intervene to modify this behavior, and improve the effectiveness of his training with this learner. However, users of e-learning platforms can easily lose their motivation and concentration. How to determine the learner style in e-learning platform, in real time? The objective of this work is to apply the research already done in the field in our Polydisciplinary Faculty of LARACH (FPL) distance learning platform using evolutionary video processing by integrating Genetics Algorithms (GA) in order to detect learner style in a video sequence.

1. Introduction

Some of us find it best and easiest to learn by listening to information, some of us by seeing them. Promoters say that teachers should assess their students' learning styles and adapt their classroom methods to better suit each student's learning style. Students who have a learning style compatible with that of the teacher and the course tend to be more successful and motivated than those who do not. As many science teachers teach in a similar style, many students are left behind [1].

The Experts in this field have proposed several solutions to resolve this problem, namely indicators and behavior scores, which can be the result of an analysis of the user's interactions with the system, and the uses of pedagogical tools, behavioral analysis system [1], which is able to group learners according to their behavior and to adapt the educational content to their needs through the use of traces to create profiles, which includes learners with the same behavior.

Thus, the problem of learner style detection is related to the development of an effective student model [6]. The researchers propose techniques for the automatic elaboration of this model using artificial intelligence methods. This later is able to support some basic activities, such as testing students' positioning, automatically correcting their productions, accompanying students in case studies or application exercises. It also creates artificial tutors or conversational agents that now make it possible to considerably increase dialogue time in foreign languages.

Artificial intelligence is increasingly being used as an alternative to more traditional techniques for developing the learner model. The techniques considered are case-based reasoning, rules-based systems, artificial neural networks, fuzzy models, genetic algorithms, multi-agent systems, reinforcement learning and hybrid systems.

The most interesting solution in our case is the uses of Eye tracking technologies and Genetic algorithm, these technologies have become one of the techniques used recently in the adaptive learning; the experiences in this domain use either eye tracking [2] or the genetic algorithm [3, 4, 5, 6, 7]. From these experiences, we can conclude that the majority of the research tends toward the use of these techniques in a system of learning.

The aim of this paper is to apply the research already done in the field in our E-learning platform of the Faculty Polydisciplinary of Larache (FPL) using evolutionary video processing by integrating Genetics Algorithms (GA) in order to detect learner style in a video sequence.

The plan adopted in this paper is as follows: the second section is devoted to the problems of learner style facing a distance learning interface and the solutions proposed by researchers in this field to solve these problems. The third section describes the work on our adopted technique (GA). The fourth one presents the research related to our proposed approach. Then, we present the Model of a Smart Learning Sequence. Finally, the last section draws conclusions and perspectives of this work.

2. Learner style

Learning style is the student's preferred learning style. Recognition is the way to improve the quality of our learning. Learning style is an innate ability based on how our brain works most effectively to obtain new information. It is not related to age, the knowledge and skills we have acquired or our intelligence. There is no desirable learning style or not. People who may have different learning styles are the most successful. It is essential to be aware of our learning style. Moreover, it is useful to know how to adjust your specific learning styles and strategies according to the learning re-sources and the learning situation. The ability to adopt your skills can improve the rate and quality of your learning.

Learner features and the needs of the learner have an important role in education. Thus, learning styles are accorded a great importance in the literature of previous times [7].

In research, there are mainly five learning style models, which are studied in the engineering education literature. These learning style models include Myers-Briggs Type Indicator (MBTI) [8]; Kolb's model [9]; Felder and Silverman learning style model [10]; Herrmann Brain Dominance Instrument [11].

To detect learner style, the expert proposes several solutions. They apply different techniques to predict learning style like the user interface tracking, and collect data from the interface, inference system or the artificial intelligent methods (Artificial Neural network, Bayesian Network, Genetic Algorithms, etc).

Among these learning styles, in our proposed work, we used models of the inheritance of genetic information, which was used in Evolutionary Video Processing.

3. Earlier studies on Genetic Algorithm

Genetic algorithms (GA) are a family of adaptive research procedures that have been described and analyzed in depth in the literature [13] [14]. GAs take their name from the fact that they are loosely based on models of genetic change within a population of person.

In GA, a population of individuals is randomly selected. These individuals are subject to several genetic operators inspired by the evolutionary in biology to produce a new population containing the best individual. This population evolves more and more until a stopping criterion is satisfied and declaring obtaining optimal best solution.

Thus, the performance of GA depends on the choice of operators, who will intervene in the production of the new populations. For each step, there are several possibilities. The choice between these various possibilities allows creating several Variants of Genetic Algorithms (VGA) to improve the GA.

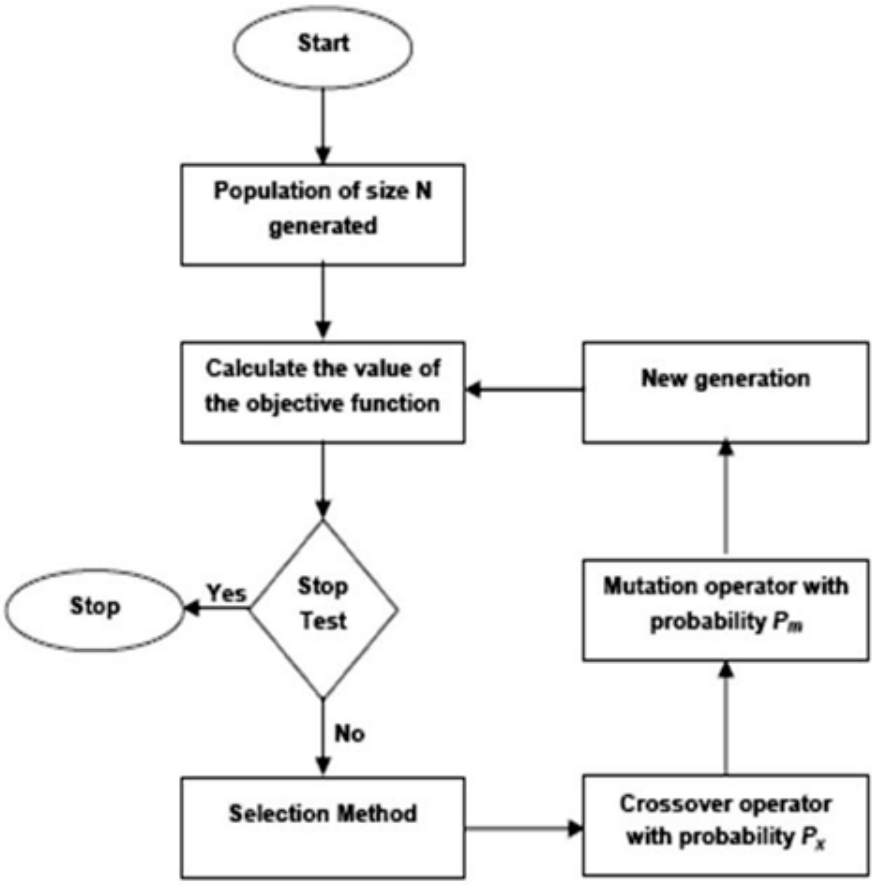

Figure 1: Genetic Algorithms

3.1. Adaptive E-learning using Genetic Algorithms

In Herrmann's work, he described an adaptive system designed in the following order to generate learning paths adapted to the learner and the pedagogical objective of the current training. He studied the problem as an "optimization problem". Using genetic algorithms, the system looks for an optimal starting path. From the learner's profile to the pedagogical objective via intermediate courses. To prepare the courses for adaptation, the application creates a description sheet for the resources, in XML while integrating it into the database [12].

3.2. AdELE: A framework for adaptive E-learning through eye tracking

The elaboration of a Meta learning system to have a self-adaptive learning system is based on the genetic algorithm to improve the parameters of self-adaptive learning. Thus, this technique has been successful for the adaptation of a scenario in a phase of teaching process by user and the construction of a tool "Course-Map" for the user to know their progress during the course [4].

4. Related works

4.1. Using genetic algorithm for eye detection and tracking in video sequence

This experiment aims to solve the problem of high tolerance in human head motion and real-time processing. For this, they proposed a method of invariant ocular tracking of size and orientation at high speed, in order to acquire numerical parameters to represent the size and orientation of the eye. For the realization of eye detection in an active scene, they proposed a matching model with a genetic algorithm [2].

4.2. The inheritance of genetic information

In case of video processing, it is very difficult to use the information between video images. Generally, in order to detect a moving object, a difference image between the images is used as information between the video images. However, it is difficult to use the image of difference in our system, because the human head moves intensively [3].

In fact, without making a new population, the ocular detection for a next image is done with the population used in a last image. Although the frame rate is 30 frames per second, it is amazing that prodigious changes occur with geometric parameters, such as location, scale and angle of rotation simultaneously to the chromosome. In other words, genetic information, which has evolved through evolution, is useful for the next frame. Therefore, this method can be expected to reduce processing time and increase accuracy [3].

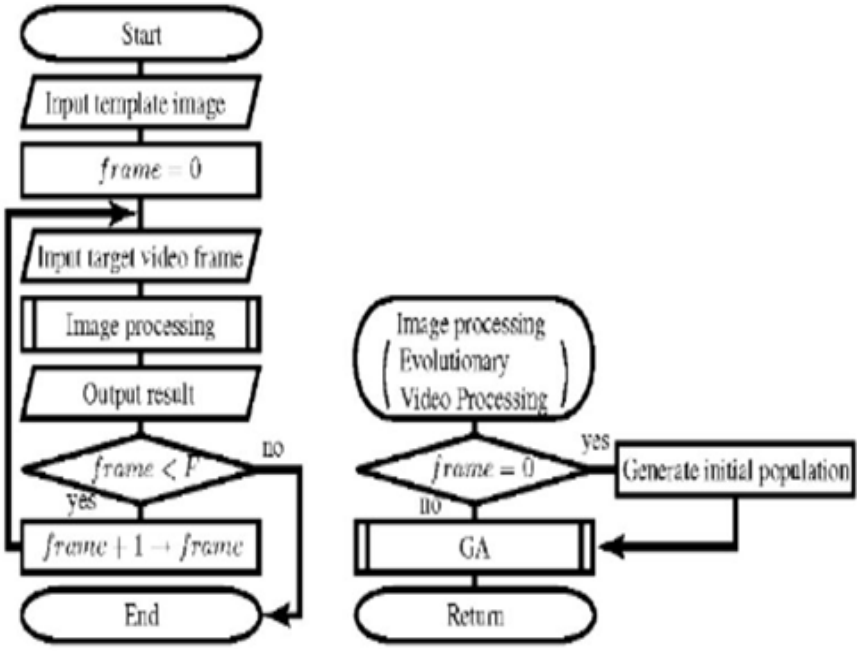

Figure 2: Flow charts: a) main process; b) evolutionary single image processing

5. Model of a smart learning sequence: proposed approach

Our approach is based on the work mentioned above which uses the genetic algorithm for eye detection in a video sequence. Our approach will be adapted for a learning process in an E-Learning system. The diagrams of our approach are shown in Figure 2. The figure represents the main part of our approach. In diagram (1), the GA process is illustrated. Diagram (2) presents the new population. In Figure 1, sequence and generation are variables that count the number of sequences in training unit and the number of generations in GA. Our approach consists of a dual runway; the outer runway (see Figure 1), and the inner runway is for GA (see Diagram

(1)). After an initial population is generated with random numbers, the GA is launched.

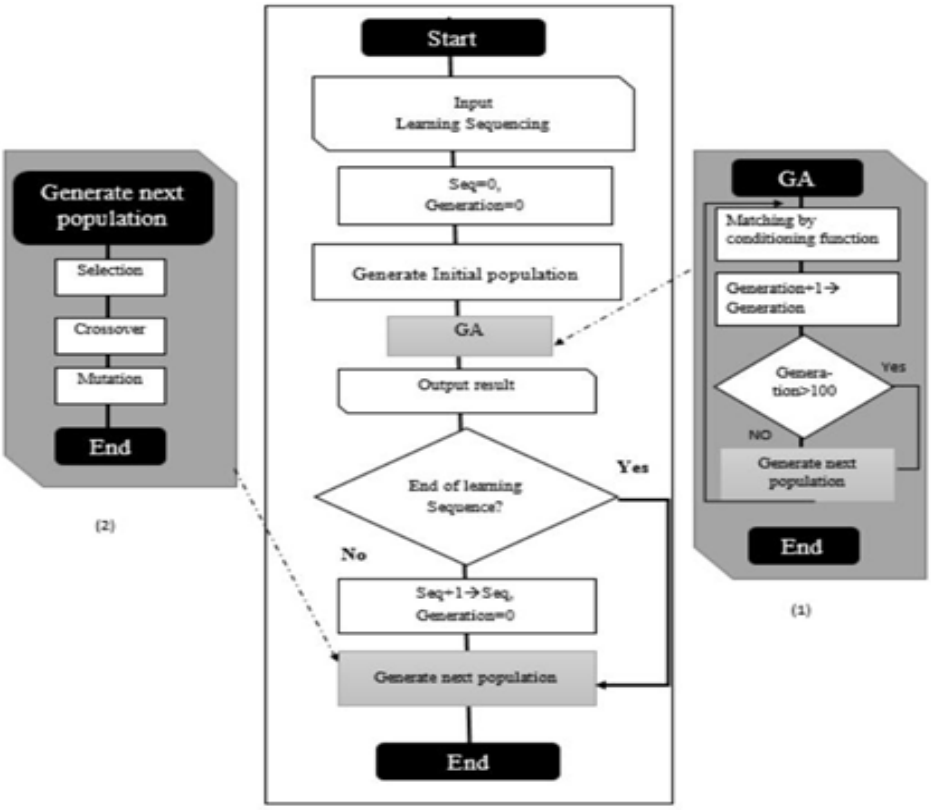

Figure 3: Model of a smart learning sequence

In this document, the GA ends if production is greater than 100. If the termination criterion is not met, a new population of the next generation is generated based on an assessment of the fitness of each individual.

In Figure 3, special attention should be given to initialization. Of the GA population only when sequence = 0. From the last GA, some genetic information from the last GA is inherited at the General Assembly. This method is described above. Through the Evolutionary Learning Sequence Processing, we can detect and track. In addition, we can extract its geometric information with great accuracy in real time.

6. Conclusion

The invariant method of detection and monitoring of size and orientation at high speed presented by the researchers, allows acquiring numerical parameters to represent the size and orientation of the eye. In this article, we discussed the fact that a high tolerance for human head movement and real-time processing is necessary for many applications, such as eye tracking. An artificial template is used for the generality of the method.

To solve these problems, we use a matching template with the genetic algorithm. A fast and reliable tracking system has been implemented by progressive video processing for ocular detection and tracking. In addition, an artificial iris template was used for the artificial iris analysis for the generality of the method [2].

Our future work will be based on the application of our proposed approaches in our FPL E-Learning system, after well on the execution of these experiments already dealt with in this article and then extract a set of criteria to classify a learner according to his style.

References

[1] Felder, R. M. (1993). Reaching the second tier - learning and teaching styles in college science education. Journal of College Science Teaching. 23(5): 286-290.

[2] Akashi, T., Wakasa, Y., Tanaka, K. (2007). Using Genetic Algorithm for Eye Detection and Tracking in Video Sequence. Systemics, Cybernet-Ics and Informatics. 5 (2):72-78.

[3] Akashi, T., Wakasa, Y., Tanaka, K., Karungaru S., Fukumi, M. (2013). Evolutionary Video Processing for Lip Tracking, International Journal of Intelligent Computing in Medical Sciences and Image Processing.

[4] Barrios, V.M.G., Gütl, C., Preis, A.M., Andrews, K., Pivec, M., Mödritscher, F., Trummer, C. (2004). AdELE: A Framework for Adaptive E-Learning through Eye Tracking, Proceedings of IKNOW.

[5] Merad, D., Metz, S.M.V., Miguet, S. (2006). Eye and gaze tracking algorithm for collaborative learning system, ICINCO-RA. Sétubal, Portugal, pp. 326-333.

[6] Py, D. (1998). Quelques Méthodes d'Intelligence Artificielle pour la Modélisation de l'Elève, Sciences et techniques éducatives. Vol. 5, No. 2.

[7] Martin, F. (1986). Phenomenography – research approach to investigating different understandings of reality. Journal of Thought. No.21, pp.28–49.

[8] Papanikolaou, K., Grigoriadou, M. (2004). Accommodating learning style characteristics in adaptive educational hypermedia. Workshop on individual differences in adaptive hypermedia organised in AH2004, Part I, pp. 77–86.

[9] Pittenger, D. J. (1993). The utility of the Myers–Briggs type indicator. Review of Educational Research. No. 63, pp. 467–488

[10] Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. NJ: Prentice Hall.

[11] Felder, R., Silverman, L. (1988). Learning and teaching styles in engineering education. Journal of Engineering Education. 78(7):674-681.

[12] Herrmann, N. (1989). The creative brain. Lake Lure, North Carolina, USA: The Ned Herrmann Group.

[13] Madani, Y., Bengourram, J., Erritali, M., Hssina, B., Birjali, M. (2017). Adaptive e-learning using Genetic Algorithm and Sentiments Analysis in a Big Data System. Vol 8, No. 51.

[14] Kenneth, D. J. (1988). Learning with genetic algorithms: An overview, Machine Learning. Vol 3, pp. 121-138.

[15] Grefenstette, J. (1986). Optimization of control parameters for genetic algorithms, IEEE Transactions on Systems Man and Cybernetics. 16(1): 122-128.

Analyse Sémantique de Conversations pour l'Identification des Profils Comportementaux sur Facebook

CHAABI Youness[1], LEKDIOUI Khadija[2], MESTADI Walid[2]

[1] CEISIC, Institut Royal de la Culture Amazighe, Rabat, Maroc

[2]LASTID, Université Ibn Tofail, Kenitra,Maroc

chaabi@ircam.ma

khadija.lekdioui@uit.ac.ma

mestadi.walid@gmail.com

Résumé

Dans cet article, nous décrivons une nouvelle approche Multi-Agents (SMA) pour l'accompagnement et le suivi d'apprenants (tutorat) dans le cadre d'apprentissage collaboratif à distance via les technologies réseaux. En effet, afin d'assister les apprenants dans leur processus d'apprentissage (de collaboration), le système que nous proposons permet de dégager un profil comportemental (sociologique) de chaque apprenant sur la base de l'analyse automatique des conversations textuelles asynchrones échangées entre ces derniers. Nous décrivons d'abord, les profils sociologiques que nous utilisons dans notre modèle. Ensuite, nous exposons l'approche utilisée pour l'analyse sémantique des messages échangés (full text), ainsi que les indicateurs proposés pour la détermination de ces profils. Nous présentons après les résultats de la mise en application du système développé dans le cadre d'une expérimentation que nous avons menée avec les étudiants du Master au sein de l'Université Ibn Tofail de Kenitra, Maroc. Les résultats obtenus lors des tests sur des corpus de messages montrent de bonnes performances.

1. Introduction

Notre travail se situe dans le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH). Nous nous intéressons dans cet article à l'Apprentissage Collectif Assisté par Ordinateur (ACAO), plus connu sous sa dénomination anglaise Computer-Supported Collaborative Learning (CSCL).

Dans le contexte d'une formation collaborative à distance, la communication textuelle est d'une importance capitale. Les outils de communication asynchrones textuels tels que le courriel et le forum évitent les contraintes bloquantes de rendez-vous. Ces outils restent à ce jour le meilleur compromis entre souplesse et interactivité pour la réalisation d'un travail collaboratif en ligne [1]. Face à la grande quantité des messages déposés dans les forums, les tuteurs se sentent souvent démunis pour construire une représentation synthétique de l'activité des individus et des groupes. Le tuteur risque de manquer d'objectivité quand il s'en sert pour évaluer l'implication et la place des apprenants dans les échanges et de repérer leurs comportements sociaux [2, 3]. Dans le cadre de l'analyse automatique des activités collaboratives des apprenants [2, 3, 4], nous proposons une approche d'analyse sémantique des conversations textuelles asynchrones entre apprenants pour déterminer leurs comportements sociaux.

Dans le cadre de formation à distance où il n'y a aucune interaction entre le tuteur et l'apprenant, les données rassemblées tendent à être plus imparfaites que celles obtenues par l'interaction en présentiel. La présence d'informations imparfaites est un facteur important qui mène souvent aux erreurs dans la détermination des comportements sociaux de l'apprenant. Cette imperfection semble due aux erreurs et aux approximations impliquées lors du recueil de l'information, partiellement en raison de la nature abstraite de la connaissance humaine et de la perte d'information résultant de sa quantification. La théorie des sous-ensembles flous [5] se présente comme un outil privilégié pour la modélisation des situations présentant des imprécisions. Une des motivations principales de l'usage de

la logique floue est la manipulation améliorée de l'imperfection de l'information. En effet, le raisonnement d'un système de logique floue est considéré comme « facile », du point de vue compréhension et/ou modification par des concepteurs et utilisateurs. Un des facteurs qui appuie cette considération est la similarité humaine. La logique floue peut fournir des descriptions de la connaissance comme un humain et imiter son modèle de raisonnement concernant les concepts vagues [6, 7]. Ceci est d'un intérêt particulier dans la conception d'un système modélisant la connaissance interprétable de l'apprenant qui est basée sur le raisonnement et la conceptualisation de l'enseignant-expert.

2. Les profils de comportements humains

Dans ses travaux en éthologie, Robert Pléty a beaucoup étudié le comportement d'élèves travaillant en groupe. Il a notamment analysé des interactions entre élèves travaillant, en groupe de quatre, à la résolution de problèmes d'algèbre [8, 2]. En nous appuyant sur ces travaux, nous avons étudié les profils de comportements sociaux dans le cadre d'un travail collaboratif en ligne. Ainsi à partir des expérimentations, nous sommes parvenus à retrouver les mêmes profils de comportement chez les étudiants travaillant en groupe sur les réseaux sociaux. Ceci en nous appuyant sur le volume d'intervention de chacun, le type de son intervention ainsi que ses réactions aux autres (ce que les interventions entraînent). C'est ainsi que nous retrouvons, dans le cadre d'environnements collaboratifs en ligne, à l'image des environnements présentiels classiques présents presque systématiquement dans tous les groupes étudiés : les profils animateur, vérificateur, quêteur et indépendant.

Profil	Volume des interventions	Type des interventions	Réactions entrainées
Animateur	Important	Fait une proposition, poste un message d'organisation et/ou encouragement. Intervient pour calmer un conflit.	Suivi de réactions
Vérificateur	Assez Important	Réagi à une proposition (Propose éventuellement). Evalue une proposition.	Suivi partiel de réactions
Quêteur	Peu Important	Ne fait pas de proposition. Pose des questions ou il exprime ses doutes sur une démarche ou proposition (esprit plutôt négatif).	Ses questions sont très bien acceptées
Indépendant	Faible	Ses interventions semblent décalées par rapport à la discussion en cours et souvent non suivies de réactions des autres membres du groupe.	Ses interventions restent en suspens

Tableau 1 : Caractéristiques des profils comportementaux d'apprenants travaillant en groupe – Adaptation à partir des travaux de [8, 3]

3. Architecture du système

L'approche proposée consiste à analyser automatiquement le contenu des messages. Suite à cette analyse, un profil pour chaque apprenant est déterminé. Le défi réside dans l'identification de patterns comportementaux des apprenants via l'analyse automatique du contenu des messages échangés. Chacun de ces messages subit une séquence de traitements. Dans cet article, nous présentons quatre profils différents que nous avons identifiés et caractérisés via différents critères. Afin de déterminer un profil, quatre traitements sont effectués. Le premier consiste à simplifier les messages en supprimant les informations inutiles. Le deuxième traitement procède à une analyse sémantique de message. Dans le troisième traitement, le système calcule les indicateurs qui seront traités par un modèle de la logique floue. Le dernier traitement permet de

déterminer un profil comportemental pour chaque acteur humain du système.

Nous mentionnons que l'interactivité entre tuteur-apprenants ou apprenant-apprenant se fait essentiellement à travers l'échange textuel. Ci-dessous l'architecture qui décrit le fonctionnement du système (Figure 1).

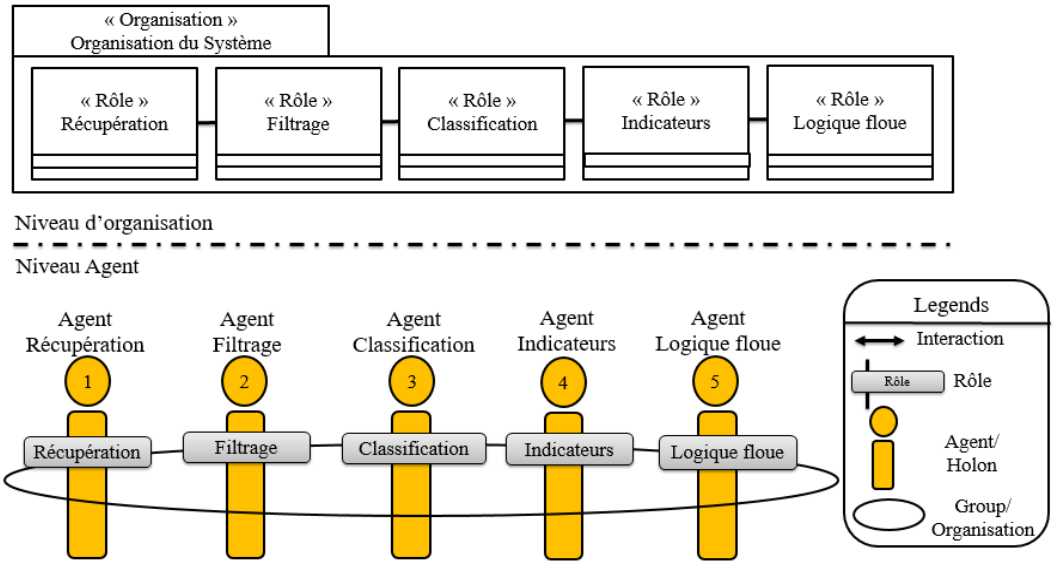

Figure 1 : Architecture du système

3.1. Agent récupération

En premier lieu, un module de récupération permet l'extraction des messages d'interaction sur les réseaux sociaux (Facebook, Twitter), ainsi que leur préparation aux différents traitements ultérieurs.

D'après notre expérimentation, un premier traitement du corpus réside dans la correction des fautes d'orthographe et de grammaire. Les erreurs d'orthographe et de grammaire peuvent se produire dans l'analyse de texte pour les humains ainsi que pour les logiciels [9, 10]. Un mot mal orthographié (ou une phrase) peut modifier complètement l'analyse.

La correction orthographique et grammaticale obtenue à l'aide d'un mot de dictionnaire (corpus) est associée à un algorithme qui prend en compte la variation de la langue (conjugaison verbale ; accord de noms et d'adjectifs). Il consiste à comparer les mots du texte avec le corpus, en tenant compte du

contexte des phrases. Malgré l'utilité du correcteur d'orthographe et de grammaire, il ne peut pas remplacer une vérification humaine attentive.

3.2. Agent filtrage

Une fois la tâche de récupération est réalisée, un second processus se charge d'effectuer le traitement de filtrage. En effet, dans les messages textuels, de nombreux mots apportent peu d'informations sur le message concerné. Nous supprimons automatiquement ces mots en utilisant des « mots vides » propres à chaque langue.

Les mots qui apparaissent le plus souvent dans un corpus sont généralement les mots grammaticaux vides de sens (mots vides) : les articles, les prépositions, les mots de liaisons, les déterminants, les adverbes, les adjectifs indéfinis, les conjonctions, les pronoms et les verbes auxiliaires, etc. Ces mots constituent une grande part des mots d'un texte, mais malheureusement sont faiblement informatifs sur le sens d'un texte puisqu'ils sont présents sur l'ensemble des textes [11].

3.4. Agent classification

L'agent classification permet de mesurer la similarité sémantique qu'un nouveau message appartienne à l'un des quatre catégories (animateur, vérificateur, quêteur et indépendant) à partir de la proportion de messages de formation appartenant à cette catégorie.

Pour commencer, nous souhaitons préciser le contexte d'extraction des messages de formation. Nous avons travaillé régulièrement avec un certain nombre de tuteurs sur des échanges entre apprenants, nous en sommes venus à aborder un ensemble des messages spécifiques à chaque catégorie des profils.

En se basant sur les propositions des tuteurs, l'analyse intuitive des messages montre que ces derniers peuvent être classés comme suit : des messages qui ont comme but d'initier une interaction et d'amorcer un sujet de discussion, des messages où on demande des informations et où on attend une réponse de la part d'autrui, des messages où on répond aux sollicitations des autres, où on répond aux questions et aux interrogations des autres et enfin des messages antérieurs qui clarifient ou approfondissent un sujet actuel de discussion.

3.4.1. Mesure de similarité topologique

Il existe essentiellement deux types d'approches qui calculent la similitude topologique entre les concepts ontologiques :

- Approche basée sur les arêtes : utilise les arêtes et leurs types comme source de données [17].
- Approche basée sur le contenu de l'information : les principales sources de données sont les nœuds et leurs propriétés [18, 19].

3.4.2. Similarité sémantique

La similarité sémantique ou la relation sémantique est un concept de mesure de la proximité entre des termes ou documents dans le contexte de leur signification. Nous avons deux méthodes différentes pour calculer la similarité sémantique. L'une consiste à définir une similarité topologique, en utilisant l'ontologie pour définir une distance entre les mots. L'autre est basée sur l'utilisation des moyens statistiques tels que le modèle d'espace vectoriel pour établir une corrélation entre les mots et les contextes textuels à partir d'un corpus de texte approprié. Nous nous concentrons sur la première approche utilisant l'ontologie WordNet pour le calcul de similarité sémantique [20]. Le calcul de similarité dans cette approche repose sur le fait que la similarité dépend des caractéristiques communes et distinctes des objets.

3.4.3. Processus de mesures de similarité sémantique

L'agent classification permet de réaliser une séquence complète de traitement. Le processus de calcul de similarité sémantique est illustré à la figure 3. Ce processus se compose de trois phases :

- Phase 1 : Module de construction temporaire
- Phase 2 : Module de calcul sémantique
- Phase 3 : Procédure de mesure de similarité sémantique

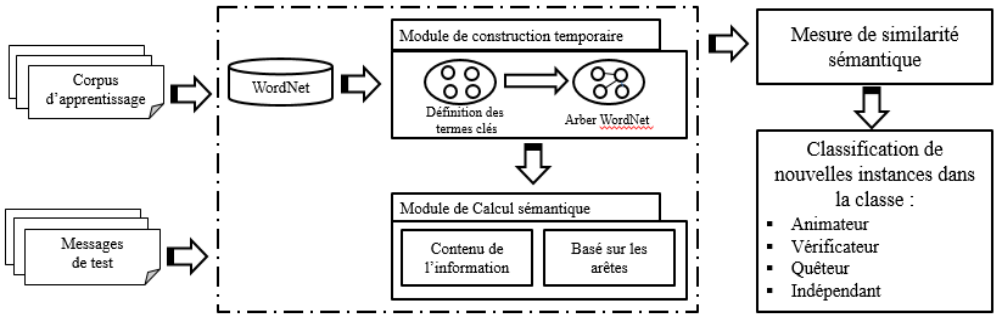

Figure 3 : Diagramme de calcul de similarité sémantique

3.4.3.1. Phase 1

L'objectif du module de construction temporaire est de sélectionner tous les mots du texte qui existent sur WordNet et d'obtenir la relation entre ces mots. Nous utilisons WordNet pour générer une représentation de texte plus riche. Dans ce module, nous avons utilisé les hyperonymes fournis par WordNet comme des fonctionnalités utiles pour l'analyse de textes.

3.4.3.2. Phase 2

Dans le module de calcul sémantique, nous utilisons les différents algorithmes qui utilisent les mesures de similarité sémantique pour trouver les sens appropriés des mots selon le contexte à l'échelle de la phrase ou du texte.

Nous citons quelques algorithmes qui permettent de calculer la similarité sémantique :

a) Longueur du chemin
b) Similarité de Resnik
c) Similarité de Lin
d) Distance de Jiang-Conrath
e) Mesure de Wu et Palmer

3.4.3.3. Phase 3

Dans les procédures de mesure de similarité sémantique, les vecteurs sémantiques pour T1 et T2 peuvent être formés à partir de T et des

statistiques de corpus. Le processus de dérivation de vecteurs sémantiques pour T1 (13) :

$$Word\ w, define$$
$$sim(w_1, w_2) = max_{c1,c2}[sim(c1, c2)]$$
$$sim(T1, T2) = \sum_{i=1}^{n} \frac{sim(Wi, Wi+1)}{n} \quad (13)$$

Nous obtenons des valeurs de mesure de similarité sémantique pour chacun des cinq algorithmes ci-dessus entre message T1 et message T2 (14):

- Chemin Sim (T1, T2) = valeur1
- Sim Resnik (T1, T2) = valeur2
- Sim Lin (T 1, T2) = valeur3
- Sim JC (T1, T2) = valeur4
- Sim Wu (T1, T2) = valeur5

$$sim(T1, T2) = Max(valeur1, valeur2, valeur3, valeur4, valeur5) \quad (14)$$

Les messages sont composés de mots, il est donc raisonnable de représenter un message en utilisant les mots qu'il contient. Contrairement aux méthodes classiques qui utilisent une liste de mots précompilée contenant des centaines de milliers de mots, notre méthode forme dynamiquement les vecteurs sémantiques uniquement sur la base des messages comparées. Les recherches récentes en analyse sémantique sont généralement adaptées pour extraire automatiquement un vecteur sémantique de mots pour une phrase [26]. Avec deux messages T1 et T2, un ensemble de mots conjoint est formé avec (15) :

$$T = T1 \cup T2$$
$$= \{W_1, W_2 \ldots, W_n\} \quad (15)$$

L'ensemble de mots T contient tous les mots distincts de T1 et T2. La morphologie flexionnelle peut amener un mot à apparaître dans un message avec des formes différentes qui portent une signification particulière pour un contexte spécifique. Pour cette raison, nous utilisons la

forme du mot tel qu'il apparaît dans le message. Par exemple, garçon et garçons, femme et femmes sont considérés comme quatre mots.

Considérons l'exemple suivant, où nous avons deux messages :

Message1 = "Marketing est les activités d'une société liée à l'achat et la vente d'un produit ou service. Il comprend la publicité, les produits, la vente et la livraison aux gens"
Message2 = "Le marketing est à la fois l'activité, l'ensemble des institutions et des processus visant à créer, communiquer, délivrer et échanger les offres qui ont de la valeur pour les clients. "

Premièrement, après le filtrage de message la matrice sera comme suit :

Message1 = {marketing, l'activité, société, associé, acheter, vendre, produit, service, compris, publicité, vendre, livrer, produit, gens}

Message2 = {marketing, activité, ensemble, institution, processus, créer, communiquer, livrer, l'échange, l'offre, valeur, client}
Deuxièmement, nous calculons la similarité sémantique entre le Message1 et Message2, qui se traduit comme suit:

Le nombre total de mots dans le Message 1 = 14
Le nombre total de mots dans le Message 2 = 12
Le nombre total de mots existants dans WordNet = 12

Le système génère 12 vecteurs sémantiques : La figure 4 représente la matrice des vecteurs sémantiques entre le Message1, le Message 2 :
Vecteur 1 sémantique = {1, 0.93, 0.37, 0.33, 0.35, 0.6, 0.25, 0.82, 0.4, 0.85, 0.6, 0.6, 0.25, 0.42}
Vecteur 2 sémantique = {0.93, 1, 0.4, 0.23, 0.37, 0.63, 0.26, 0.87, 0, 0.26, 0.63,0, 0,26, 0.54}

.....

Jusqu'au vecteur sémantique 12

	Marketing	Activité	Société	Associé	Acheter	Vendre	Produit	Service	Compris	Publicité	Vendre	Livrer	Produit	Gens	Max
Marketing	1	0.93	0.37	0.33	0.35	0.6	0.25	0.82	0.4	0.85	0.6	0.6	0.25	0.42	1
Activité	0.93	1	0.4	0.23	0.37	0.63	0.26	0.87	--	0.26	0.63	--	0.26	0.54	1
Ensemble	0.42	0.46	0.57	0.33	0.4	0.33	0.28	0.4	0.4	0.2	0.33	0.33	0.22	0.72	0.72
Institution	0.4	0.42	0.93	0.23	0.37	0.31	0.26	0.37	--	0.54	0.31		0.26	0.66	0.93
Processus	0.88	0.93	0.37	0.29	0.35	0.6	0.25	0.82	0.33	0.63	0.35	0.22	0.25	0.46	0.93
Créer	0.4	--	0.29	0.4	0.4	0.3	--	--	0.5	--	0.3	0.47	--	--	0.5
Communiquer	0.25	--	0.2	0.25	0.25	0.22	--	--	0.28	--	0.22	0.18	--	--	0.28
Livrer	0.23	--	--	0.6	0.44	0.86	--	--	0.5	--	0.86	1	--	--	1
Echange	0.33	--	--	0.33	0.5	0.85	--	--	0.4	--	0.85	0.57	--	--	0.85
Offre	0.43	0.46	0.42	--	--	--	0.46	0.4		0.76	--	--	0.46	0.54	0.54
Valeur	0.42	0.44	0.31	0.33	0.33	--	0.21	0.4	--	0.35	--	--	0.21	0.37	0.44
Client	0.22	0.22	0.21	0.76	--	--	0.53	0.4	--	0.27	--	--	0.53	0.3	0.76

Figure 4 : Matrice sémantique entre message 1 et message 2

A titre d'exemple la similarité sémantique entre « Marketing » et « Activité » est égale à 0.93.

La colonne Max (Figure 4) représente la valeur max de vecteur de sémantique = {1, 1, 0.72, 0.93, 0.93, 0.5, 0.28, 1, 0.85, 0.54, 0.44, 0.76}.

D'aprés le tableau 2, nous obtenons la similarité sémantique entre le Message1 et Message2 : est égal à 0,76 :

Algorithmes	Chemin	Resnik	Lin	Jiang	Wu	Max
Valeur Sémantique	0.47	0.70	0.73	0.74	0.76	0.76

Tableau 2 : Résultats de calculs sémantiques par les algorithmes longueur du chemin, Resnik, Lin, Jiang, Wu et Palmer

3.5. Agent indicateur

A partir des données recueillies par le système, nous calculons trois types d'indicateurs [6, 8] qui permettent de définir des profils sociaux des apprenants. Les indicateurs peuvent être classés selon deux natures :

quantitative (le volume d'interventions et réactions entrainées) et qualitative (type d'interventions).

3.5.1. Volume d'interventions

Une première formule permet de calculer tout d'abord le ratio de participation d'un apprenant (P) en divisant le nombre de messages envoyés par celui-ci par le nombre de messages envoyés par les apprenants du même groupe (X) (16).

$$RI = \frac{nbrMsgApprenant(P)}{NbrTotaleMessagesGroupe(X)} * 100 \qquad (16)$$

3.5.2. Type d'intervention

Après l'analyse sémantique et la classification des conversations, quatre expressions permettent de calculer le type d'intervention pour identifier la catégorie de profil : les messages de catégorie Animateur, Vérificateur, Quêteur et Indépendant. Le type d'intervention d'un apprenant P est calculé comme suit (17) :

$$Animateur = \frac{nNbrMessagesAnimateur}{NbrTotaleMessages(A,V,Q,I)} * 100 \qquad (17)$$

Avec NbrTotaleMessages (A,V,Q,I) est Le nombre total des messages des catégories Animateur, Vérificateur, Quêteur et Indépendant de l'apprenant P.

Les calculs des autres types d'interventions (Vérificateur, Quêteur, Indépendant) sont obtenus de manière similaire.

3.5.3. Réactions entrainées

Selon les caractéristiques des profils définis (Tableau 1), le volume de réactions entrainées par un message permet de caractériser un profil comportemental. A titre d'exemple, un profil Animateur exige un suivi de réactions très important par rapport à celui d'un Vérificateur.

Une conversation est vue comme une co-construction dans laquelle les interlocuteurs interagissent de manière cohérente. D'une manière

générale, nous retenons que « chaque interaction répond à une intervention précédente et en même temps impose des contraintes discursives sur l'interaction suivante ». Plus précisément, les sujets qui ont un grand nombre de réponses et interactions sont des sujets d'intérêt pour le groupe. Nous calculons alors pour chaque message les réactions directes (1ère réaction à un message) et les réactions indirectes (nombre des interventions faites après le lancement du message).

Réaction directe : Les réponses directes sur les messages d'un apprenant (18).

$$Reaction_{Directe} = \frac{\sum_{i=1}^{m} reponse\ sur\ \ Message_i(App)}{NbrReponseTotale(Groupe\ X)} * 100 \qquad (18)$$

Réaction indirecte : profondeur de la discussion (19).

$$Reaction_{Indirecte} = \frac{TA}{\sum_{i=1}^{g} TA_i} * 100 \qquad (19)$$

Avec :

$$TA = \sum_{i=1}^{m} Profondeur - \sum_{i=1}^{m} Reponse\ directe$$

m : Nombre des messages envoyés par l'apprenant.

g : Nombre d'apprenants dans un groupe.

3.6. Agent logique floue

La plupart des problèmes rencontrés sont modélisables mathématiquement. Mais ces modèles nécessitent des hypothèses parfois trop restrictives rendant délicate leur application au monde réel. Les problèmes du monde réel doivent tenir compte d'informations imprécises et incertaines. Les connaissances dont disposent les humains sur le monde ne sont presque jamais parfaites. Elles sont presque toujours entachées d'une quantité d'incertitudes et d'imprécisions. Nous ne parlons pas ici du raisonnement scientifique, dont l'objectif est justement de se débarrasser de toute imperfection, mais de tous les autres raisonnements que nous faisons tous les jours, sans cesse, sur les choses, les personnes et les pensées nous environnent [5].

La logique floue semble donc reproduire la flexibilité du raisonnement humain quant à sa prise en compte des imperfections des données accessibles. Il serait donc intéressant de l'utiliser au cœur des systèmes experts, systèmes dont le but est de reproduire les mécanismes cognitifs d'un expert dans un domaine particulier. La logique floue peut également servir pour un système décisionnel lors de la phase d'analyse des données par exemple. Elle peut s'avérer utile pour la prise de décision, soit pour découvrir des règles ou inférences floues permettant de mieux comprendre les données et ainsi éclairer les décisions, soit pour effectuer des requêtes dites floues en se basant sur les connaissances des experts.

En effet, l'algorithme flou se déroule en 3 étapes :

- transformation de variables quantitatives en variables logiques floues ;
- utilisation de règles logiques pour évaluer de nouvelles variables floues en sortie ;
- transformation de ces variables floues en variables qualitatives.

Finalement, notre système sera doté d'une estimation qualitative des comportements sociaux de l'apprenant (Figure 9), lui permettant ainsi d'identifier ses lacunes et ses points faibles et d'équilibrer les groupes en fonction de leurs comportements sociaux.

Figure 9 : Centres de gravité profil Animateur, Vérificateur, Quêteur et Indépendant

4. Contexte général de l'expérimentation

Le modèle objet de ce travail a pour finalité d'automatiser certaines tâches (laborieuses) habituellement effectuées par un tuteur humain. Dans ce sens, nous avons procédé à une étude comparative entre l'évaluation humaine et celle produite : résultat de notre modèle.

Nous avons mené des expériences d'analyse intuitive sur des conversations d'apprenants. Nous nous intéressons ici à l'analyse qualitative et quantitative réalisées auprès de 3 tuteurs. Un corpus de messages fut élaboré à partir d'un échantillon remis par 9 groupes de 4 apprenants, sur une période de 4 mois (du 02 mars au 02 juin), organisée en 6 phases, correspondant à des tâches différentes. Notre corpus est constitué de 80 à 120 messages échangés par les élèves au sein de leur même groupe pour chaque phase du projet. Cette analyse de conversations textuelles se base sur les caractéristiques des profils comportementaux définis plus haut (Tableau 1). L'analyse intuitive des messages consiste à : (1) associer un profil à chaque message, (2) identifier les actes de langage qui déterminent le profil et (3) associer un profil à chaque étudiant. Dans le cadre d'une pédagogie par projet par exemple, ces observations fournissent à l'enseignant encadrant des indications pour comprendre, réagir et intervenir auprès du groupe. De même, pour les apprenants, la perception des comportements des individus du groupe permet de mieux réguler le travail collectif.

Les figures 10 et 11 illustrent les résultats de l'analyse intuitive faite par les tuteurs pendant les deux premières phases du projet, dont l'objectif est d'associer un profil comportemental à chaque apprenant. Pour plus de lisibilité, dans cette analyse, chaque encadrant va associer un profil à chaque apprenant selon son comportement social. En effet, après avoir identifié les profils apprenants, nous avons demandé aux tuteurs de faire (par analyse des contenus) une classification des messages (type : animateur, vérificateur, quêteur, indépendant) en identifiant les actes de langage qui les caractérisent : proposition, message d'organisation et/ou encouragement, intervention pour calmer un conflit, réaction à une proposition, expression de doutes sur une démarche ou proposition, etc. (voir tableau 1). A partir des résultats d'analyse des tuteurs pour le groupe 1 durant les périodes 1 et 2 (Figure 10 et 11), l'impression des profils dégagée par les apprenants, durant chaque période du projet, est confirmée par les données relevées par le système à travers l'analyse de contenus des

messages (Figure 12 et 13). Par exemple, pour le groupe 1 durant la période de rédaction du cahier des charges, l'apprenant 1 dégage un profil animateur qui correspond à l'analyse des données relevées par le tuteur : volume d'intervention élevé (44,32%) et type d'intervention élevé.

	Groupe	Impression de profil dégagée par l'apprenant	Type des interventions				Volume d'intervention
		Profil	Nombre Message Catégorie Animateur	Nombre Message Catégorie Vérificateur	Nombre Message Catégorie Quêteur	Nombre Message Catégorie Indépendant	
Apprenant 1	Groupe 1	Animateur	15 34,88 %	21 48,83 %	7 16,27 %	0 00,00 %	Important 44,32 %
Apprenant 2		Animateur	11 47,82 %	11 47,82 %	1 04,34 %	0 00,00 %	Important 23,71 %
Apprenant 3		Indépendant	3 50,00 %	3 50,00 %	0 00,00 %	0 00,00 %	Faible 06,10 %
Apprenant 4		Animateur	13 52,00 %	11 44,00 %	1 04,00 %	0 00,00 %	Important 25,77 %

Figure 10 : Résultat de l'analyse intuitive pour le groupe 1dans la phase 1 du projet

	Groupe	Impression de profil dégagée par l'apprenant	Type des interventions				Volume d'intervention
		Profil	Nombre Message Catégorie Animateur	Nombre Message Catégorie Vérificateur	Nombre Message Catégorie Quêteur	Nombre Message Catégorie Indépendant	
Apprenant 1	Groupe 1	Animateur	22 53,65 %	12 29,26 %	5 12,19 %	2 04,87 %	Important 41,41 %
Apprenant 2		Vérificateur	3 15,78 %	13 68,42 %	2 10,52 %	1 05,28 %	Assez Important 19,19 %
Apprenant 3		Indépendant	1 16,66 %	3 50,00 %	0 00,00 %	2 33,33 %	Faible 06,06 %
Apprenant 4		Animateur	16 48,48 %	12 36,00 %	3 09,09 %	2 08,06 %	Important 33,83 %

Figure 11 : Résultat de l'analyse intuitive pour le groupe 1 dans la phase 2 du projet

Lorsque nous soumettons les mêmes données d'interaction entre apprenants au système d'analyse automatique que nous proposons, nous

obtenons les résultats exposés dans les figures 12 et 13, pour le même groupe et pour les mêmes périodes.

	Résultats Analyse Système pour le Groupe N°1 Phase Rédaction Cahier de Charge							
	Volume d'intervention	Type des interventions				Réactions entrainées		Profil
		Animateur	Vérificateur	Quêteur	Indépendant	Directe	Indirecte	
Apprenant 1	44,33 %	35,20 %	30,54 %	15,62 %	00,00 %	36,53 %	38,70 %	Animateur
Apprenant 2	23,71 %	44,50 %	42,00 %	04,00 %	01,23 %	26,92 %	35,23 %	Animateur
Apprenant 3	06,19 %	04,80 %	40,00 %	00,00 %	03,25 %	04,80 %	09,25 %	Indépendant
Apprenant 4	25,77 %	48,84 %	43,69 %	03,38 %	06,50 %	28,84 %	39,50 %	Animateur

Figure 12 : Indicateurs calculés par le système pour le groupe 1 dans la phase 1 du projet

	Résultats d'analyse du système pour le Groupe N°1 Phase Rédaction Cahier de Charge							
	Volume d'intervention	Type des interventions				Réactions entrainées		Profil
		Animateur	Vérificateur	Quêteur	Indépendant	Directe	Indirecte	
Apprenant 1	44,33 %	35,20 %	30,54 %	15,62 %	00,00 %	36,53 %	38,70 %	Animateur
Apprenant 2	23,71 %	44,50 %	42,00 %	04,00 %	01,23 %	26,92 %	35,23 %	Animateur
Apprenant 3	06,19 %	04,80 %	40,00 %	00,00 %	03,25 %	04,80 %	09,25 %	Indépendant
Apprenant 4	25,77 %	48,84 %	43,69 %	03,38 %	06,50 %	28,84 %	39,50 %	Animateur

Figure 13 : Indicateurs calculés par le système pour le groupe 1 dans la phase 2 du projet

L'analyse de ces résultats, à la lumière des caractéristiques des profils d'apprenants définis (Tableau 1), permet d'associer un profil sociologique à chaque apprenant. Vu les résultats des analyses sémantiques pour calculer le type d'intervention, deux profils se dégagent : Animateur et Vérificateur. Cependant, en analysant les volumes d'interventions et réactions entrainées associés, nous constatons qu'ils sont importants et caractérisent le profil Animateur. Ainsi, pour l'exemple considéré et pour la période 1, l'apprenant (1) sera qualifié d'« Animateur ». La même approche a été

utilisée pour définir les profils apprenants dans les périodes 1 et 2 pour les étudiants de groupe 1 (Figures 12 et 13).

Nous avons pu apprécier, à travers cette étude, l'utilité de la notion de la logique floue dans l'évaluation des niveaux comportementaux de l'apprenant. La théorie des sous-ensembles flous procure une méthode adéquate pour incorporer la connaissance d'un enseignant expert en utilisant des termes qualitatifs et proches du raisonnement humain. Elle permet ainsi de manipuler des informations imprécises et de modéliser des connaissances subjectives. De plus, l'utilisation des règles floues dans l'algorithme d'inférence du système fournit à l'utilisateur plus de souplesse et de facilité dans son processus de jugement. Par ailleurs, nous avons représenté l'évolution dans le temps des profils (Animateur, Vérificateur, Quêteur et Indépendant) d'un apprenant (1) présent sur les groupes de discussions et sur les réseaux sociaux (Figure 14).

Par exemple, le graphe signale que l'apprenant (1) a majoritairement tenu un rôle animateur pendant la première phase du projet. Par ailleurs, nous pouvons remarquer que cet apprenant est devenu Vérificateur à la deuxième phase.

Cette vue permet d'identifier le rôle joué par les apprenants dans leur groupe à travers les différentes phases du projet. Cette variation est le résultat de préférences chez les apprenants par rapports aux tâches associées à chaque phase du projet.

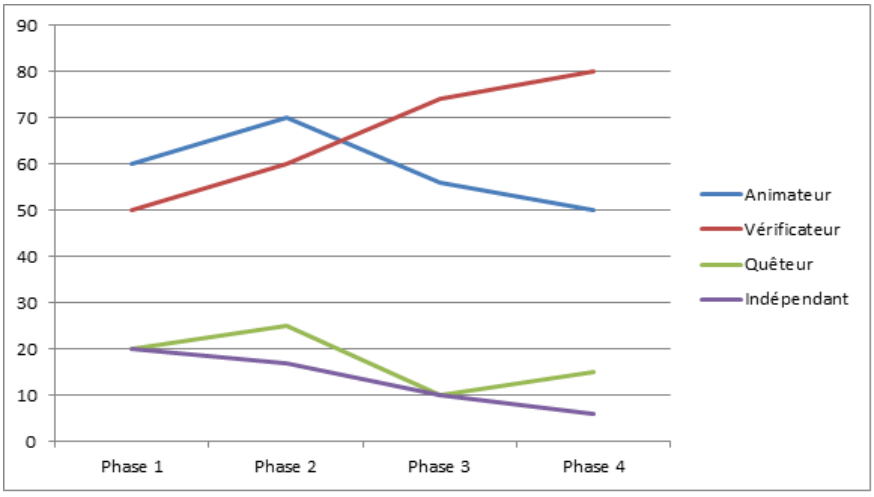

Figure 14 : Evolution du profil de l'apprenant (1) sur toute la durée du projet

Vu ces résultats, nous constatons que les observations faites par les tuteurs sont confirmées par l'analyse automatique faite par notre système, ce qui nous amène à affirmer que notre approche permet de retrouver les profils comportementaux dans les groupes d'apprenants travaillant à distance.

5. Conclusion

Le présent travail s'inscrit dans le cadre du développement de systèmes d'Analyse Automatique des Interactions (AAI), très largement utilisés pour répondre aux contraintes posées par le tutorat à distance via les technologies réseaux. Nous proposons une procédure complète d'analyse « fulltext » des échanges textuels pour la détermination de profils sociologiques d'apprenants dans le cadre des processus de formation collaboratifs à distance. Cette analyse consiste en 2 étapes (Récupération et Filtrage) au bout desquelles, nous effectuons une analyse sémantique de conversations qui va par la suite contribuer à la classification des messages (type Animateur, Vérificateur, Quêteur et Indépendant). En se basant sur les profils définis et adaptés à partir des travaux de Pléty [2], nous calculons des indicateurs qui, couplés aux classifications de messages décrites ci-dessus, permettent d'affecter un profil sociologique à chaque apprenant sur la base de logique floue. L'approche a été testée sur une situation réelle qui a montré une grande concordance entre les résultats observés par des tuteurs humains et ceux déterminés automatiquement par notre système. Comme perspective au développement de ce projet, nous prévoyons d'intégrer un système de recommandation qui permet de déclencher des mécanismes d'alerte en direction des tuteurs et/ou des apprenants.

Références

[1] Jaillet, A. (2011). Le tutorat en formation à distance, De Boeck, Coll. Perspectives en éducation et formation.

[2] Pléty, R. (1998). Comment apprendre et se former en groupe. Retz pédagogie.

[3] George, O., Mathern, B., Mille A., Bellet, T., Bonnard, A., Henning N., Trémaux, J.-M. (2007). Analysis of Behavior and Situation for mental Representation Assessment and Cognitive activity modelling, Retrieved April 10.

[4] Sujana, J., Claire, M., John, K. (2012). A visualisation tool to aid exploration of students' interactions in asynchronous online communication. Computers & Education. 58(1):30-42.

[5] Zadeh, Z. (1965). Fuzzy sets, Information and Control. 8(3):338–353.

[6] Serin, S., Morova, N., Saltan, M., Terzi, S., Karaşahin, M. (2014). Prediction of the marshall stability of reinforced asphalt concrete with steel fiber using fuzzy logic. Journal of Intelligent & Fuzzy Systems. 26(4):1943-1950.

[7] Karakasidis, T. E., Georgiou, D. N. (2004). Partitioning elements of the Periodic Table via fuzzy clustering technique. Soft Computing. 8(3):231-236.

[8] Pléty, R. (1996). L'apprentissage coopérant. Presses Universitaires Lyon.

[9] Hantler, S. L., Laker, M. M., Lenchner, J., Milch, D. (2017). Methods and apparatus for performing spelling corrections using one or more variant hash tables. U.S. Patent No. 9,552,349. Washington, DC: U.S. Patent and Trademark Office.

[10] Rashmi, S., Hanumanthappa, M. (2017). Qualitative and quantitative study of syntactic structure: a grammar checker using part of speech tags. International Journal of Information Technology. pp. 1-8.

[11] Rakholia, R. M., Saini, J. R. (2017). A Rule-Based Approach to Identify Stop Words for Gujarati Language. In Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, Springer, Singapore. pp. 797-806.

[12] Wu, Z., Palmer, M. (1994). Verbs semantics and lexical selection. In Proceedings of the 32nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics. pp. 133-138.

[13] Zhu, G., Iglesias, C. A. (2017). Computing Semantic Similarity of Concepts in Knowledge Graphs. IEEE Transactions on Knowledge and Data Engineering. 29(1):72-85.

[14] Hua, W., Wang, Z., Wang, H., Zheng, K., Zhou, X. (2017). Understand Short Texts by Harvesting and Analyzing Semantic Knowledge. IEEE transactions on Knowledge and data Engineering. 29(3):499-512.

[15] Jiang, J. J., Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy, arXiv preprint cmp-lg/9709008.

[16] Pedersen, T., Pakhomov, S. V., Patwardhan, S., Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. Journal of biomedical informatics. 40(3):288-299.

[17] Girardi, D., Wartner, S., Halmerbauer, G., Ehrenmüller, M., Kosorus, H., Dreiseitl, S. (2016). Using concept hierarchies to improve calculation of patient similarity. Journal of Biomedical Informatics. Vol 63, pp. 66-73.

[18] Wang, H. Y., Ma, W. Y. (2017). Integrating Semantic Knowledge into Lexical Embeddings Based on Information Content Measurement. EACL 2017. pp. 509.

[19] Hao, W. U., Huang, H. (2017). Efficient Algorithm for Sentence Information Content Computing in Semantic Hierarchical Network.

[20] Meštrović, A., Calì, A. (2016). An ontology-based approach to information retrieval. In Semanitic Keyword-based Search on Structured Data Sources. Springer, Cham. pp. 150-156

[21] Sudha, S. (2016). WordNet–An On-line Lexical Database.

[22] Rada, R., Mili, H., Bicknell, E., Blettner, M. (1989). Development and application of a metric on semantic nets. IEEE transactions on systems, man, and cybernetics, 19(1):17-30.

[23] Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, Canada. pp. 448–453.

[24] Lin, D. (1998). Automatic retrieval and clustering of similar words. In Proceedings of the 17th international conference on Computational linguistics. V. 2, pp. 768-774. Association for Computational Linguistics.

[25] Wu, Z., Palmer, M. (1994). Verb semantic and Lexical Selection. In Proceddings of 32 Annual Meeting of the association of computer Linguistics (ACL 994), Las Cruces, New Mexico.

[26] Navarro-Almanza, R., Licea, G., Juárez-Ramírez, R., Mendoza, O. (2017). Semantic Capture Analysis in Word Embedding Vectors Using Convolutional Neural Network. In World Conference on Information Systems and Technologies. pp. 106-114.

Building Virtual Worlds for Amazigh Game Based Learning

TAZOUTI Yassine[1], BOULAKNADEL Siham[2], FAKHRI Youssef [1]

[1] Laboratoire LaRIT Faculté des Sciences, Université Ibn Tofail,Kénitra
[2] Institut Royal de la Culture Amazighe, IRCAM Rabat
yassine.tazouti@gmail.com
boulaknadel@ircam.ma
fakhri@uit.ac.ma

Abstract

In this paper, we present ImALeG (Immersive Amazigh Learning Game), a game to learn Amazigh language. ImALeG is an ongoing serious language game designed for interactively learning Amazigh language vocabulary. The game makes use of virtual reality environment provided by unity platform to implement immersive learning, and multi agent concept, learn-by-doing techniques that have proved to be along with the most efficient learning strategies to comprehend language. We have focused on the learning of Tifinaghe alphabet as a first step that we consider essential before learning vocabulary.

1. Introduction

Nowadays the situation of Amazigh language in Morocco have changed. Its status reconsidered, being official language besides Arabic. This factor makes teaching of the Amazigh language more prevalent not only in primary schools, but also in some Moroccan universities. Therefore, thinking of sophisticated tools allowing learning easily rather than traditional methods becomes a necessity [3]. Today, the new generation is expert in using technology. We can say learning Amazigh by gaming would be one of the greatest methods of this group of people. It is motivating, available anytime and anywhere. In other hand off-school

situation is an important feature to teach concepts through serious games to learners. Beside it, current educational systems have their own disadvantages like transport, i.e., traffic, get free for travelling, consequently education ministries might consider distance and eLearning system. Moreover, the Moroccan learners in Amazigh classroom spend majority of their time copying texts on their notebooks, doing more exercises and preparing lessons every night so that his schooling in elementary school will be as a permanent job punishment and not as a productive learning. However, many studies proved that when learners play, they develop their skills in many ways. They think, solve problems, express themselves, move, cooperate, use their impressions and exercise their moral conscience [4]. From this point of view, we can say that serious game provides a good opportunity for learning and it could be an alternative for present education. In this paper, we present ImALeG, a language game designed for interactively learning Amazigh language. The game uses the 3D virtual reality environment exploiting 'learn-by-doing' techniques for language acquisition. The 3D environment allows learning process personalization and enhances learner awareness of the acquired learning skills. Furthermore, it permits ubiquitous learning as well as universal access, features that aid the collection of big amounts of test data over the web and make the game excellent as a tool for comparing different teaching methodologies.

2. Related work

Many research studies have examined the potential of serious games and showed that exploiting them as educational tools might be efficient. According to [1, 2, 3], computer games proved to be promising in the learning field, as they can offer some factors that the traditional teaching methods lack. In the last decade, the learning language research has been interested on applying 3D virtual reality environments, like a Second Life, for languages' acquisition. Indeed, this type of environments affords immersive learning and facilitates distance learning. In addition, these technologies could promote learning as a social experience providing learners to use active communication over the web by means of chatting, emails exchange. In this context, many projects have emerged, in the last few years, such as Zengo Sayu system, which was developed in 1996 by Howard Rose at the University of Washington's Human Interface Technology Laboratory [6]. It uses a Virtual Reality tool, which allows to

the learner to discover a room while selecting and moving objects. The user learns by listening to audio cues from the system based on the objects they interact with and their location in reference to other objects in the environment. A further system, entitled SGLL ProjectX system, was built in 2008 at the Dublin Institute of Technology's School of Computing. The ProjectX system puts users in a 3D environment with a static camera and allows the users to move around by selecting objects in the scene, which the camera changes to focus on. The system is basically a set of mini-games related to a common theme [7]. The user, for example, had the ability to gather ingredients to make a recipe in a 3D kitchen environment or carrying out a grocery list in a 3D marketplace environment. An online system, known as Middworld Online, was developed in 2010 by Middlebury College's Interactive Languages division. This system offers to the users an online interactive world allowing them immersed in the language through coursework, activities, and mini-games [8]. One such mini-game places users in charge of waiting tables at a restaurant. The users have the ability to interact with customers and maintain them all happy by determining what they ordered and how to answer to them correctly in the target language. Making errors will result in unhappy customers and a lower overall score. In the other hand, there are many studies that exploit multi-agent systems as an adaptation technique in serious games, due to their behavior and dynamic ability to change the global game settings. Therefore, there have been a few notable language learning projects also situated in multi-agent systems. One of these projects, which offers design architecture of serious online games based on monitoring the characters to follow the development of the player's skills to guarantee a real-time adaptation of the agent's behaviors, and to avoid scenario disturbance during adaptation. The project has developed an organizational structure for planning agents to follow the storyline progression [9]. Another system was developed, in 2007, by Delmas. It has a dynamic generation of gaming scenario through a multiagent architecture that supervises the player's reaction time, interactions and preferences through a linear programming to calculate the parameters of scenario game such as the execution time and the game play from these information collecting multi-agent system deducts the possible states of the player using a petri network [11]. This work is based on the theory of emergent narrative. It consists to build a dramatic story of the game during its execution.

Generally, this approach has been used in training Therapeutic games. The architecture of the proposed multi-agent system developed by Alvarez is called ALIVE [12]. This architecture is designed to produce a non-centralized control method and improvement of nonplayer characters behaviors in order to become more real. This enhancement allows specifying behavior of nonplayer characters on a principle of "why the action must be done" and not "how it will be done". This adaptation technique also allows specifying metrics used to follow and evaluate agents' behavior.

3. Amazigh language

3.1. Approaches to Amazigh language teaching

Amazigh is taught most intensively in Moroccan primary schools, usually as a timetabled subject. In both independent and state sectors, Amazigh is more expected to be an enrichment or extra option than a main timetable subject. The most usually used approaches to Amazigh language teaching take account of the Grammar Translation Method (GTM), the Audiolingual Method (ALM) and Communicative Language Teaching (CLT). Nevertheless, numerous researchers [13; 14; 15] claim that rather than assuming a single perspective for language learning, aspects from a range of approaches should be adopted and the four skills (speaking, listening, reading and writing) must be well balanced within the classroom. The Grammar Translation Method (GTM) was for a long time the main language teaching pedagogy for Amazigh. The GTM concerns linguistic form as being the primary object of teaching and learning. The approach was adapted to assist learners read literature rather than to develop conversation skills in the language. Actually, the Communicative Language Teaching (CLT) is the approach adopted by Moroccan primary school for teaching Amazigh language. It is inspired from the pedagogy of skills. The textbook designers assume that the teaching of Amazigh aim at mastering the communicative competence of the learner, mainly oral skill that is based on dialogue and secondarily written skill, both in terms of expression and comprehension. Gradually, once acquired the strategies of oral, reading and writing, the learner learns cultural competence through activities, these pedagogical approaches seems to be traditional and does not apply to learners of this generation who tend to explore new information technologies that are so far unused in our educational system.

From this point, we have found that serious games will be useful for Amazigh teaching as a pedagogical medium or an e-learning support that would have an important advantage [5], based on the motivation of learners through an emerging narrative story and transmission skills using feedbacks through learning by trial and error approach. Furthermore, to stimulate interaction, we integrate a NPC to present the appropriate aids for each learning situation for achieving goals and accomplishing missions.

3.2. Common Amazigh learning difficulties

The most often-cited reason for labeling Amazigh as a difficult language is simply because it contains rather little in common with the European language families. Belonging to the Afro-asiatic language family, Amazigh has had a 3 path of development particularly different from the big European languages. In learning Amazigh language, the learners suffer from enumerable issues: 1) In reading, such as: the lack of punctual marks, the inability of distinguishing similar letters, – the inability of decoding graphemes and pronouncing uncommon words. 2) In spelling, teachers assume that spelling influences the reading and the writing performance, and improves vocabulary and comprehension skills. In other words, learning to spell helps the connection between the graphemes and their sounds. Amazigh spelling is hard but when the learner masters its morphological, syntactical and semantic structures, it is perfectly decodable. 3) In grammar, since Amazigh was oral tradition, its grammar is more irregular and has a lot more patterns than English or French. 4) Amazigh morphology is considered rich and complex in terms of its inflections involving infixation, prefixation and suffixation.

3.3. Amazigh graphical system

Like any language passes throw oral to a written mode, the Amazigh language has required a graphical system. In this aspect, there were in Morocco three graphical systems for Amazigh. These systems are those of Arabic, widely used for religion and rural poetry writing; Latin supported by the International Phonetic Alphabet (IPA), used particularly by berberists since early works of missionaries; and Tifinaghe, the ancestral writing system, which has been preserved by Touareg. In Morocco, the choice ultimately fell on Tifinaghe for technical, historical and symbolic reasons. Since the Royal declaration on February, 11th 2003, Tifinagh has become an official graphic system for writing Amazigh, particularly in

schools. Thus, the IRCAM has developed an alphabet system called Tifinaghe IRCAM. This alphabet is based on a graphic system towards phonological tendency. This system does not retain all the phonetic realizations produced, but only those that are functional. The Tifinaghe-IRCAM alphabet contains 27 consonants, 2 semi-consonants and 4 vowels.

3.2.1. Directionality

Historically, in ancient inscriptions Amazigh language was written horizontally from left to right and from right to left; vertically upwards and downwards; or in boustrophedon. However, the orientation most often adopted in Amazigh language script is horizontal and from left to right, which is also adopted by IRCAM.

3.2.2. Tifinaghe encoding

Over the last several years, the encoding of scripts in Unicode/ISO 106461 for rendering complex scripts has advanced rapidly. However, encoding the Tifinaghe script was not a straight prospect as one might hope— partly owing to the many variants of Tifinaghe used in different parts of the Amazigh world. The encoding is composed of four Tifinaghe character subsets: the basic set of IRCAM, the extended IRCAM set, other NeoTifinaghe letters in use, and modern Touareg letters. The two first subsets constitute the sets of characters chosen by IRCAM. While, the first is used to arrange the orthography of different Moroccan Amazigh dialects, the second subset is used for historical and scientific use. The letters are classified in accordance with the order specified by IRCAM. Other Neo-Tifinaghe and Touareg letters are interspersed according to their pronunciation. Thus, the UTC accepts the 55 Tifinaghe characters for encoding in the range U+2D30...U+2D65, U+2D6F, with Tifinaghe block at U+2D30...U+2D7F [9].

4. Immersive Amazigh language learning game

ImALeG is a serious game ongoing project developed by our group to learn Amazigh language. This scientific project is a first attempt to make a very fantastic game for learning the Amazigh language. It aims to learn vocabulary (on some specific topics as furniture, animals, food). In this paper, we have focused on the learning of Tifinaghe alphabet as a first step that we consider essential before learning vocabulary. In what follows, we

start by describing the game scenario used for language learning. We, then, describe the features, the goals and the coverage of this application, the game screen and the system architecture.

4.1. Game scenario

We imagined the game as a village populated with houses, a restaurant, a cinema, a supermarket, a zoo, and a forest. Each place of them is representing a game unit corresponding to a particular lexical field, to be acquired by the learner. The ImALeG forest is currently the main game unit to learn tifinaghe alphabet. A first-person perspective is employed whereby the learner is an avatar that can easily move in the game world. In order to start the game, the learner is prompted to fill an information form corresponding to his age, level of education, native language and sex. In the aim to define its profile and to store it in the database for later retrieval. Once these options have been set, the learner arrives at the main menu where there are a few other useful settings such as camera and audio options. The game scenario of ImALeG for learning Amazigh alphabet called tifinaghe. It consists on a forest full of tree where each tree hangs a tifinaghe alphabet. The learner is represented as an avatar and can navigate freely and interact with the virtual world by touching, moving or taking objects.

4.2. Game features

The language game showed in this paper has the following features:

4.2.1. Immersive learning

The use of the 3D environment provided by unity engine allows ImALeG to provide immersive learning. In the game, the flow of the learning practice is generated by the learner, and explicitly by his position and by the actions (s)he achieves (e.g. which letter (s)he touches) in the game world.

Figure 1: Game Environments

4.2.2. Personalization

In ImALeG, the learning process is adapted to learner needs and language proficiency. In each game session, the learner decides the content by touching specific alphabets and the type of activity he prefers to train. After each game session, the interactions of the learner with the system are saved in a database. The content of the database is extracted each time the same learner log in the game.

4.2.3. Scoring

In addition to scoring, which helps to" keep" the learner playing, the learner must full the health bar, then he will have the necessary energy that gives the right to be teleported to another level. The scoring of the different learning activities is employed to build the learner model and to establish the skill level of the learner in supplementary interactions.

4.2.4. Evaluation Data

The game has been achieved within the unity engine environment. The goal for selecting the unity engine environment is twofold: First, it offers high level 3D graphical tools to help the design of virtual worlds. Second, being available by everyone over the Internet, unity engine lets collecting large amounts of test data for evaluation.

4.2.5. Multi-agent concept

To guarantee flexibility and autonomy of the system as well as a decentralized architecture [10], we have chosen to create a multi-agent system composed of three agents, each agent has a specific behavior. The three agents are always enabled during the game session, each agent executes a specifics function, and they have in common a main mission of gathering data related to the evolution of the learner, treating and evaluating them to provide an adaptive interface. For this, we have defined the following agents: Tutor agent that is responsible for the nonplayer character movement control and the detection of learning situations that reach dialog agent. Collect Agent, which is responsible of data collection related to the player evolution in the virtual environment (scores obtained, the correct answers, errors), the transmission of this information to the Player database to keep tracking the learning evolution. Dialogue Agent that is responsible of loading dialogues' data in necessary information for the player and providing them to other agents or to the environments.

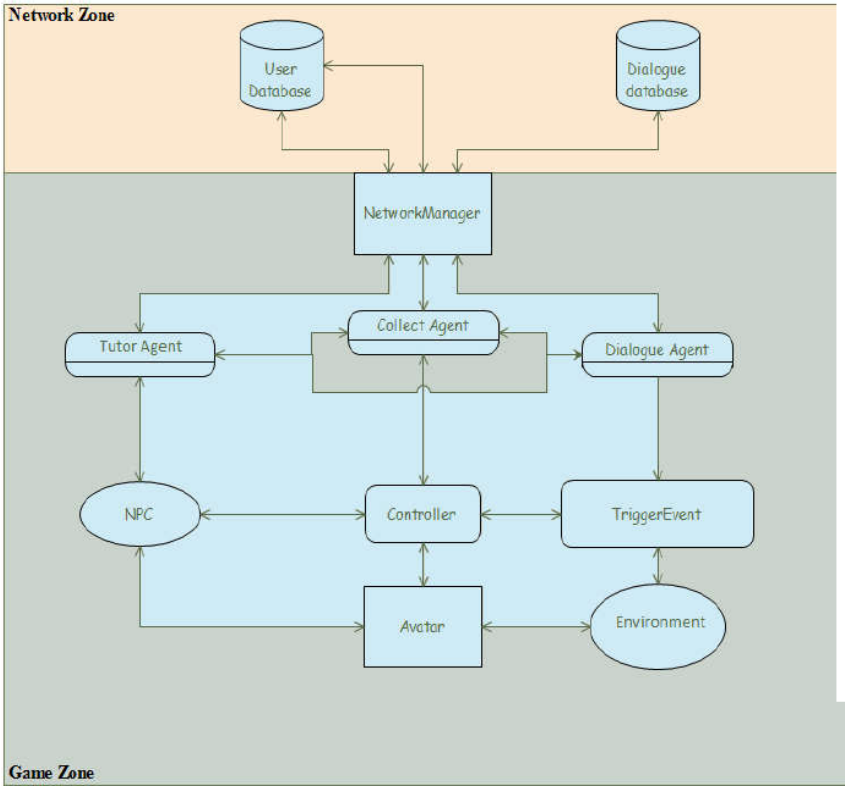

Figure 2: Multi-agent concept

4.3. Goals and coverage

The educational goal of ImALeG is to facilitate the acquisition of the Amazigh language. Currently, it can be viewed as a "situated textbook" providing the learner with alphabet drills. The current implementation targets' learner of Amazigh is at beginner's level. Each drill in the game has different levels that represent the four learning stages needed for the learner to learn a certain letter in the alphabet. According to the Montessorri syllabus, these stages can be translated into four different levels as follows: –Level 1: matching two identical objects and their names. –Level 2: recognizing a word of an object given its image. –Level 3: matching the letter with the words it starts with. –Level 4: matching the letter with the images of the objects it starts with.

4.4. Screen management

The state machine or the screen manager is intended to coordinate different "game screens." Thus, in our platform game, we have a main menu screen, a game screen, an explored worlds screen, pause screen, avatar selection screen and setting screen. In figure 3, we have created a diagram that shows all the screen and connections between them.

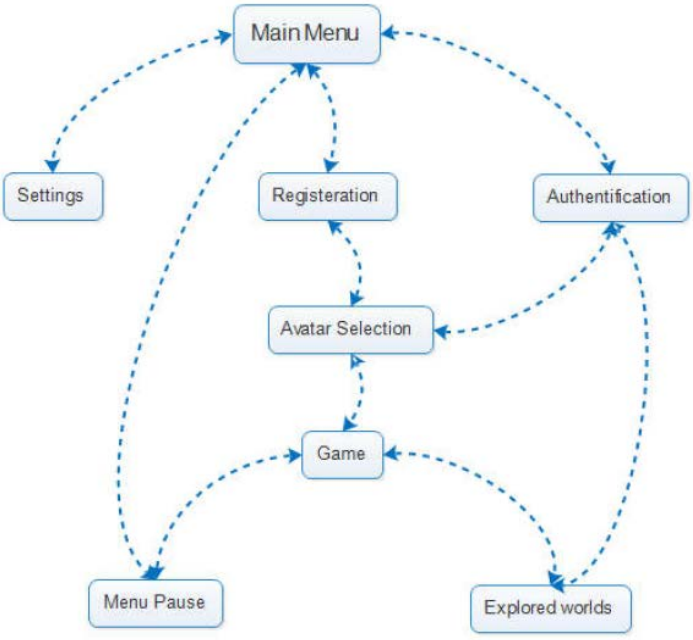

Figure 3: Diagram of screen management

4.5. Game architecture

Our game communicates with the databases using network manager of the game engine that offers the player the possibility to become a server and client at the same time and send collected information in real time through an API System. Collect agent handles the collection and calculation of scores and information that arrives from the game scene and other agents, and stores the results information corresponding to the score obtained, the number of errors made in each level and the number of feedbacks that took place. The tutor agent allows control over the nonplayer character (NPC

tutor) and loading and display conversation that comes from the dialogue agent. The implementation of a control and behavior process of a NPC is complicated in development due to the huge number of iteration. To overcome this problematic, we have used an open sources tools NavMesh Agent and RAIN AI. These two tools are behavior models to simplify and customize designing our NPC behavior. They are compatible with the integration of traditional animation. The dialogue Agent aims to detect events from loading the corresponding dialogue and information, in order to send them to others agents or to the environments.

Figure 4: Screen shot game scene

Figure 5: Game architecture

4.6. Teaching activities

In the explorer mode, the system teaches the learner basic knowledge about Amazigh alphabet. This happens at two levels:

1) When the learner enters the forest, the system indicates to her/him the name of the place in which (s)he is in. The learner can move freely in the forest. Each tree gives the learner information about itself.

2) When the learner touches alphabets in the tree, the system tells the learner their pronunciation. Double-clicking an alphabet displays the "Info Menu" screen, providing more details about the alphabet, such as the pronunciation description, translation into the learner's native language and its category either consonants, semi-consonants or vowels. The learner can touch whatever objects (s)he wants thereby obtaining information on the object.

Teaching is depending on which alphabet the learner touches in the teaching content. Thus, the learning flow will be different. For instance, the learner can decide to start learning alphabet. Farther, teaching can be set. The system can output for the alphabets touched by the learner descriptions of different complexity. The complexity of the description depends on the level of language proficiency of the learner and on the teaching goal. For instance, after the learner has touched the alphabet "a",

the system outputs his pronunciation and his writing if the learner is a beginner. However, if the learner is more advanced and the teaching goal is to teach vowels and consonants the system outputs the types of the alphabet (semi-consonant, vowel, or consonant).

5. Learning activities

After the explorer mode, all types of training activities can be activated. The game contains five activities. Each activity addresses a learning goal. The learning goals are: memory, organization, visualization, audition and writing.

5.1. Memory exercises

In this activity, the learner is inquired to combine two identical cards that are flipped down on the floor after seeking the whole panel for a short span of time. The learning goal that handles is memory learning alphabet, because it makes the learner practice remembering the letters and their corresponding words, which boosts her/his short-term memory performance.

5.2. Organization exercises

This activity is a puzzle game, where the learner has to rearrange the dispersed parts of the letter's image alongside with the images of the objects that start with this letter. The learning goal that targets is the organization alphabet learning and testing the ability of putting letters at their place.

5.3. Visualization exercises

This exercise is based on drag and drop activity. It involves a collection of empty cards with Amazigh words or letters, and below them a set of images that the learner is asked to drag and drop in the correct slot. The learning goal that handles is matching the letter with the images of the objects it starts with.

5.4. Audition exercises

In this activity, the learner crosses a path, which there is group of Amazigh alphabets. Whenever, the learner is approaching a group of tifinaghe alphabets at a specified distance, a pronunciation is triggered and the learner is prompted to recognize the alphabet and fill it in a table. The learning goal that handles is the audition alphabet learning.

5.5. Writing exercises

This activity focuses on the writing skills of the learner by having the Amazigh letter on the screen, and inquiring the learners to color it. They must not go outside the letter nor fulfil a small part of it. The learning goal that handles is the writing. This activity is special as it is apart points at teaching the letter's shape to the learner. The levels count on making the job harder for the learner. This is carried out by making the letter thinner for the learner to paint and asking a higher percentage of accuracy. It is designed for the writing learning goal, because it depends on supporting practice in writing, following directions and knowing the shapes of the letters.

6. Conclusion

In this paper, we presented the prototype of ImALeG that is an interactive 3D language game for learning tifinaghe alphabet of Amazigh language. Our game makes use of virtual reality technology allowing immersive language learning and exploits multiagent system to provide autonomy. In future work, we want to extend ImALeG by taking into account vocabulary language. We further plan to formalize the process of evaluation of the learner output, and to define more proficient behavior to present feedback to learners.

References

[1] Hulusic, V., Pistoljevic, N. (2012). LeFCA: Learning Framework for Children with Autism. Procedia Computer Science. Vol. 15, pp. 4-16.

[2] Hanus, M., Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. Computers & Education. 80, pp. 152-161

[3] Lu, A., Chan, S., Cai, Y., Huang, L., Nay, Z. T., Goei, S. L. (2017). Learning through VR gaming with virtual pink dolphins for children with ASD. Interactive Learning Environments. pp.1- 12.

[4] Yedri, O. B., El Aachak, L., Belahbib, A., Zili, H., Bouhorma, M. (2017). Learners' Motivation Analysis in Serious Games. In Proceedings of the Mediterranean Symposium on Smart City Applications, Springer. pp. 710-723.

[5] Casañ-Pitarch, R. (2018). Smartphone Serious Games as a Motivating Resource in the LSP Classroom. Journal of Foreign Language Education and Technology. Vol. 3, N° 2.

[6] Rose, H., Billinghurst, M. (2008). Zengo Sayu: An Immersive Educational Environment for Learning Japanese. Human Interface Technology Laboratory, University of Washington, Seattle, WA.

[7] M. Dunne. SGLL Project X. Dublin Institute of Technology, Dublin, Ireland.

[8] Vogel, T., Cook, S. (2010). MiddWorld Online. Middlebury College, Middlebury, VT.

[9] Westra, J., Dignum, F., Dignum. V. (2011). Guiding user adaptation in serious games, In Agents for games and simulations II, Springer Berlin Heidelberg. pp. 117-131.

[10] Bratman, E. M. (1987). Intentions, Plans, and Practical Reason. Harvard University Press, Cambridge, MA.

[11] Delmas, G., Champagnat, R., Augeraud. M. (2007). Plot monitoring for interactive narrative games. In Proceedings of the international conference on Advances in computer entertainment technology. pp. 17-20.

[12] Alvarez-Napagao, S., Koch, F., Gómez-Sebastià, I., VázquezSalceda, J. (2011). Making games alive: an organizational approach, In Agents for games and simulations II, Springer Berlin Heidelberg. pp. 179-191.

[13] Fotos. S. (2013). Traditional and grammar translation methods for second language teaching, in Hinkel, E. (ed.) Handbook of Research in Second Language Teaching and Learning, New York: Routledge. pp. 653-670.

[14] Hinke, E. (2013). Handbook of Research in Second Language Teaching and Learning, New York: Routledge.

[15] VanPatten, B. (1998). Cognitive characteristics of adult second language learners, in Byrnes, H. (ed.) Learning Foreign and Second Languages: Perspectives in Research and Scholarship (Teaching Languages, Literatures, and Cultures), New York: The Modern Language Association of America. pp. 105-127.

Conception d'un Système d'Identification de Stratégie d'Apprentissage à Base des Représentations Analogiques

MESTADI Walid[1], CHAABI Youness[2], NAFIL Khalid[3]

[1] LASTID, Université Ibn Tofail, Kenitra, Maroc

[2] CEISIC, Institut Royal de la Culture Amazighe, Rabat, Maroc

[3] Équipe de recherche en gestion de projet logiciel, Université Mohammed V, ENSIAS, Rabat, Maroc.

{Mestadi.walid, knafil} @gmail.com,

chaabi@ircam.ma

Résumé

L'apprentissage joue un rôle très important dans l'évolution des connaissances humaines. Dans l'enseignement, les enseignants constatent de plus en plus que le niveau des apprenants se dégrade et qu'ils n'arrivent pas à réutiliser leurs connaissances d'une manière efficace dans des nouvelles situations, tandis que les apprenants réclament qu'ils n'ont pas eu un apprentissage de qualité ou ils ont mal compris le cours. Devant cette situation, il serait judicieux d'examiner les mécanismes entrepris par l'enseignant pour faciliter l'apprentissage ainsi que les stratégies de transfert des connaissances chez l'apprenant.

Dans cet article, nous proposons trois niveaux de structuration du contenu à enseigner par l'élaboration des stratégies d'apprentissage et d'évaluation dans le but de faciliter la compréhension et de s'assurer des résultats d'apprentissage. La structuration des connaissances et les stratégies d'apprentissage ainsi que l'évaluation peuvent être utilisées dans des systèmes d'apprentissage numériques ou classiques.

1. Introduction

Le processus d'apprentissage joue un rôle très important pour le développement du cursus des apprenants et l'évolution des connaissances des êtres humains. Plusieurs recherches essayent d'améliorer ce processus pour assurer le besoin des entreprises en compétence. Un autre élément important qui joue le rôle du moteur de ce processus est l'évaluation. Ainsi, les informations récoltées de cette dernière permettent de prendre des décisions d'apprentissage, de valider et d'examiner le suivi de progression.

Les compétences de l'enseignant, les capacités de l'apprenant, la structuration des connaissances et la pédagogie sont des composantes qui participent à ce processus d'apprentissage. Ces composantes sont complémentaires, avec une forte relation entre elles. Ignorer une composante, peut influencer le résultat d'apprentissage et réduire sa qualité.

Le succès de ce système dépend de la motivation des apprenants et des enseignants. Ainsi, l'apprentissage classique a adopté la technologie pour profiter des bénéfices de cette dernière, en proposant des systèmes d'apprentissages numériques accessibles par le Web et des outils numériques pour solliciter la motivation des apprenants ou des personnes qui souhaitent apprendre. Ces systèmes numériques doivent garder toutes les capacités de l'enseignant et doivent être conçus comme un enseignant numérique (figure 1) qui se caractérise par :

- La spécification de plusieurs pédagogies.
- L'implémentation d'une intelligence artificielle.
- L'utilisation par les personnes qui n'ont pas la possibilité d'avoir un enseignant ou/et les sociétés pour l'évaluation des candidats et la formation de ses employés.

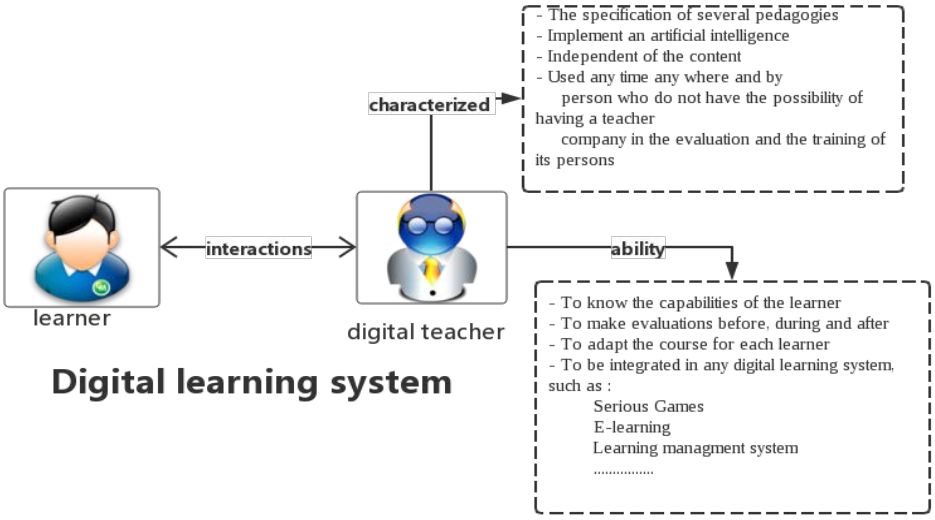
- The specification of several pedagogies
- Implement an artificial intelligence
- Independent of the content
- Used any time any where and by
person who do not have the possibility of having a teacher
company in the evaluation and the training of its persons
characterized
interactions
ability
learner
digital teacher
Digital learning system
- To know the capabilities of the learner
- To make evaluations before, during and after
- To adapt the course for each learner
- To be integrated in any digital learning system, such as :
Serious Games
E-learning
Learning managment system
.....

Figure 1 : Perspective des systèmes d'apprentissage numériques

Le développement de ce système n'est pas évident, car il doit être capable : (1) d'identifier les compétences et les acquis de l'apprenant, (2) structurer les connaissances (3) proposer des évaluations avant, en cours et après la formation, (4) personnaliser l'apprentissage tout au long de la formation selon le profil de l'apprenant.

Afin de cerner notre travail nous avons défini la problématique en deux volets :

a) Premièrement, les recherches dans le domaine d'identification de stratégie d'apprentissage à base des représentations analogiques proposent des approches difficiles à utiliser par les enseignants. Hayatie [1] a mentionné que les connaissances à enseigner, les systèmes d'évaluation ainsi que les modèles pédagogiques doivent être :

- normalisés ou unifiés dans un modèle simple.
- structurés dans un système définissant leurs interactions.

b) Deuxièmement, les apprenants mémorisent seulement le contenu du cours au lieu de le comprendre, ce qui met en question la validation des évaluations passées. Tardif [2] a mentionné "*les professeurs de sciences physiques, de sciences naturelles et de sciences humaines craignent que leurs*

élèves, par exemple, lorsqu'ils doivent recourir au graphisme, ne puissent pas réutiliser les connaissances qu'ils ont acquises en classe de mathématiques".

Lorsque l'évaluation porte sur la mémorisation, les apprenants ayant une forte capacité de mémoriser recevront de bonnes notes, par contre, les apprenants ayant une forte capacité d'analyse peuvent ne pas avoir de bonnes notes. Ainsi, le résultat peut ne pas refléter la performance des apprenants et induire à une mauvaise validation des connaissances souhaitées. De plus, les apprenants arrivant à faire le transfert de connaissances sont ceux qui ont compris le contenu souhaité, ce qui est confirmé par les travaux de Mayer [3].

Notre objectif est de concevoir un modèle permettant de contribuer au processus d'apprentissage, d'être proche aux pratiques des enseignants [1], d'être intégré dans les systèmes d'apprentissage classiques ou numériques, et aussi de proposer plusieurs stratégies d'apprentissages et d'évaluations.

2. Etat de l'art

La taxonomie de bloom permet la classification des niveaux d'acquisition des connaissances par rapport aux fonctions cognitives, elle permet à un enseignant de définir les objectifs pédagogiques pour évaluer une fonction cognitive donnée. Cette taxonomie est structurée en six niveaux (figure 2). Dans chaque niveau, elle définit les actions qui peuvent être demandées aux apprenants, notamment dans le premier niveau « mémoriser », on trouve l'action « identifier ». Chaque niveau supérieur peut avoir des actions des niveaux inférieurs. Par exemple, pour comprendre un concept, il faut l'identifier en premier lieu, dans la fonction « comprendre » est un niveau supérieur que la fonction « mémoriser ». Cependant, la taxonomie de bloom a été révisée dans les travaux de Krathwohl [7] où la dimension cognitive a été dissociée de la dimension connaissance.

Le niveau d'abstraction de cette taxonomie est indépendant du contenu à enseigner, par exemple, dans le but d'évaluer le degré de mémorisation, on peut exposer des éléments à une personne et par la suite lui demander de les identifier, de manière similaire pour les paroles d'une chanson, d'une formule mathématique, d'une scène théâtrale, etc.

The Knowledge Dimension	1. *Remember*	2. *Understand*	3. *Apply*	4. *Analyze*	5. *Evaluate*	6. *Create*
A. *Factual Knowledge*	Objective 1					Objective 3
B. *Conceptual Knowledge*		Objective 2			Objective 4	Objective 3
C. *Procedural Knowledge*						
D. *Metacognitive Knowledge*						

Tableau 1 : Processus de la dimension cognitive.

Évaluer les capacités cognitives est souvent utilisé par les entreprises pour recruter les candidats en analogie avec les besoins du poste proposé, par exemple, si le poste demande une forte capacité d'analyse, ils proposent aux candidats des évaluations qui mesurent cette fonction.

Dans les travaux de Paquette [5], ils ont proposé un langage de conception des connaissances par objets typés (figure 3), dont les types désignent qu'une connaissance peut être un concept, un fait, une procédure ou un principe, similaire à la dimension connaissance présentée par Krathwohl [7]. En outre, pour donner une sémantique aux différents types de connaissances, ils ont présenté des liens permettant de définir la relation existante entre les types de connaissances en déterminant les règles de liaison de ces types. Dans la figure 3, le caractère « c » présente le lien de composition.

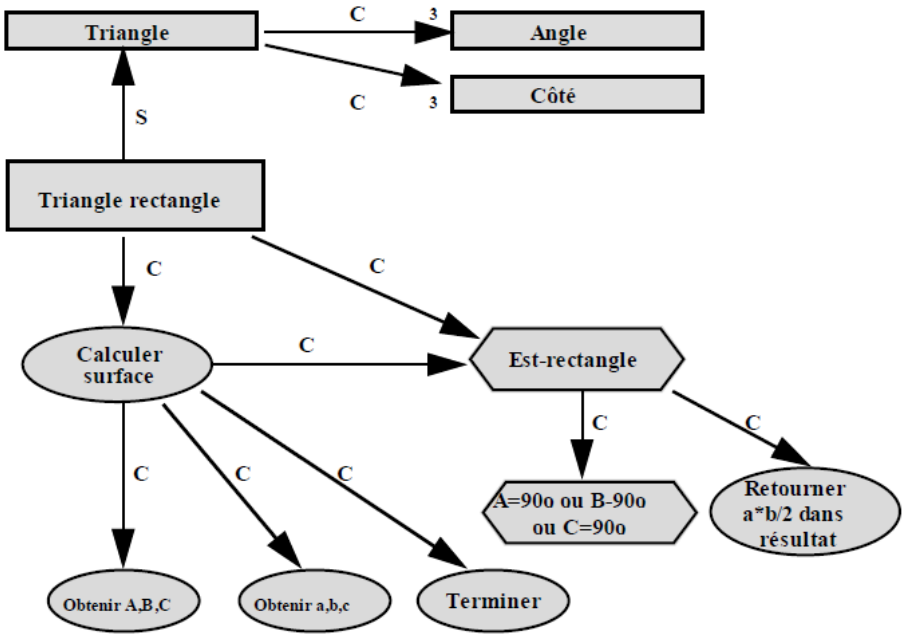

Figure 2 : Les types de connaissances et la relation de composition

La modélisation des connaissances seules n'est pas suffisante pour faciliter l'apprentissage et améliorer le transfert des connaissances. Dans les travaux de Sloutsky [4], ils ont expérimenté l'utilisation des représentations visuelles abstraites et concrètes des connaissances dans le domaine électronique. Ainsi, ils ont conclu que l'apprentissage par des représentations abstraites peut renforcer le transfert de connaissances et les représentations concrètes facilitent l'apprentissage. Dans les travaux de Moreno [6], ils ont effectué une comparaison entre l'utilisation des représentations abstraites et concrètes ainsi que la combinaison de ces deux dernières. Ils ont conclu que la combinaison des deux représentations peut améliorer l'apprentissage. L'avantage de l'utilisation des représentations concrètes peut réduire les contraintes liées à la complexité du domaine [8] et ainsi améliorer l'apprentissage par la réduction de la charge cognitive [8]. Le passage d'une représentation visuelle abstraite vers une concrète est un passage depuis une connaissance d'un domaine à enseigner dans une forme abstraite vers une instance de cette connaissance projetée soit dans le même domaine, soit dans d'autres domaines. En occurrence, Paquette [5] représente un triangle comme un niveau abstrait d'un triangle rectangulaire (voir figure 3). Tandis que Sloutsky [4] représente dans des

schémas électroniques des ampoules et des batteries au lieu de symboles abstraits.

Par contre, dans le tableau 1, les projections sont représentées dans des domaines différents (c'est-à-dire depuis l'informatique vers des situations réelles). Nous nous sommes intéressés par cette dernière. De plus, nous utilisons le terme projection pour désigner le passage d'une représentation vers une autre, spécifiquement d'un domaine vers d'autres domaines.

3. Trois niveaux de structuration des connaissances

Le processus de structuration des connaissances ne pouvait être observé que sur des indices repérables au cours des entretiens à l'intérieur des interactions dans le couple tuteur-élève. Ces indices devaient révéler la nature du traitement des connaissances mises en jeu. Les sciences de la pédagogie nous donnent des grilles utiles pour caractériser les niveaux de structuration des connaissances. Nous caractérisons trois niveaux : (1) Constatation, (2) Identification et (3) Normalisation.

3.1 Constatation

Dans un processus d'apprentissage, les enseignants ont souvent recours à des projections dans des domaines plus faciles pour expliquer une connaissance compliquée. Ainsi, le fonctionnement du cœur humain peut être assimilé à une pompe, un condensateur en électronique ou un routeur en réseau informatique.

La qualité des projections dépend fortement du niveau de maîtrise des connaissances de l'enseignant. Si la projection est mal formulée, elle peut avoir des effets négatifs sur la mémorisation et la compréhension des réciproques. Ainsi, la figure 3 présente un extrait d'un livre sur les designs patterns où l'auteur [9] a adopté la projection du contenu pour une finalité de faciliter l'apprentissage. Le label (1) représente le contenu à enseigner, (2) et (3) représente des projections dans des activités quotidiennes.

1

Le pattern Commande encapsule une requête comme un objet, autorisant ainsi le paramétrage des clients par différentes requêtes, files d'attente et récapitulatifs de requêtes, et de plus, permettant la réversibilité des opérations.

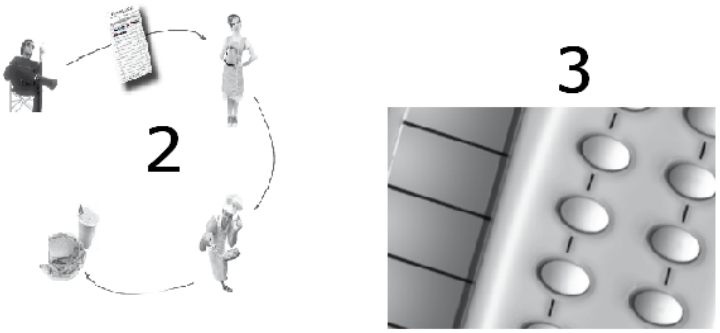

Figure 3 : Extrait depuis le livre sur les designs patterns [9].

Nous constatons que les éléments du contenu à enseigner sont habillés par d'autres éléments. Les entités dans (1) ne sont pas les mêmes dans (2) et (3), mais le point commun entre (1), (2) et (3) est qu'elles représentent le même concept.

3.2 Identification

Afin d'identifier les mécanismes permettant de projeter une connaissance, nous représentons par les logigrammes la connaissance sous une forme abstraite pour définir son aspect dynamique. Dans la suite, deux scénarios seront représentés pour comprendre les projections.

- Description du premier scénario-1 : la phase d'extraction des besoins est critique pour la conception de logiciels. D'où l'élaboration du scénario suivant :
    - Le client exprime ses exigences et valide le cahier des charges.
    - L'analyste prépare la réunion, questionne le client, analyse, décompose, classifie les exigences, et produit le cahier des charges.

- L'analyste prépare les questions à poser au client après avoir fixé un rendez-vous.
- Le client et l'analyste discutent sur les besoins lors de la réunion.
- L'analyste analyse les besoins et produit le cahier des charges qui sera validé par le client.

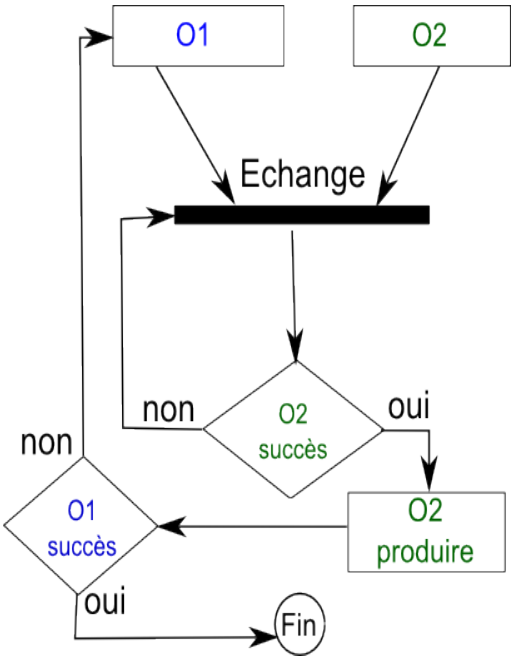

Figure 5 : Logigramme du concept d'expression de besoins après isolation du concept depuis son contexte. O1 est le client et O2 est l'analyste.

À partir de la représentation de ce scénario (figure 5), l'enseignant peut proposer plusieurs projections exprimant le même scénario. Les figures 6 et 7 illustrent respectivement une projection dans le domaine du commerce et dans le domaine de la programmation logicielle.

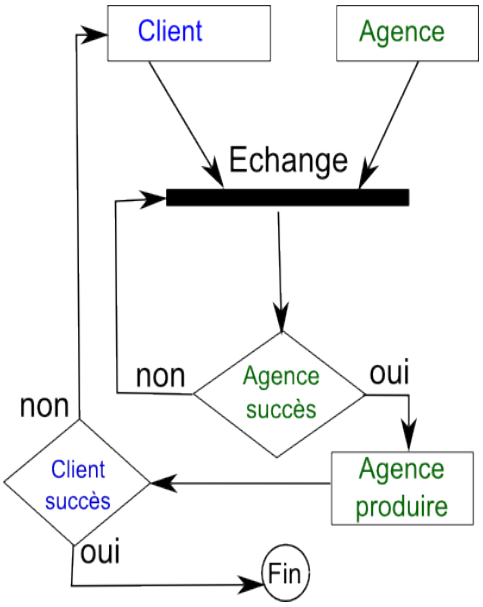

Figure 6. Logigramme du concept d'expression de besoins, dans le domaine commercial.

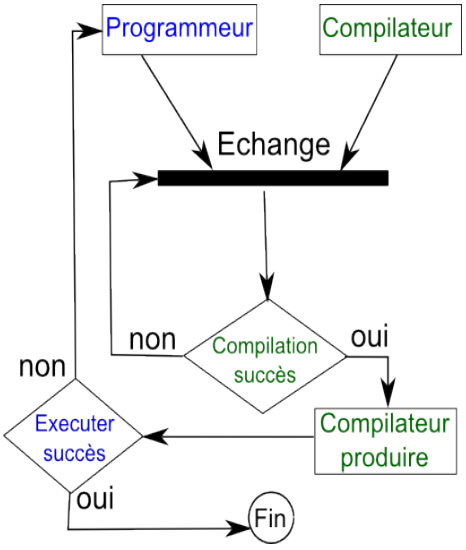

Figure 7 : Logigramme du concept d'expression de besoins, dans le domaine de la programmation logicielle.

- Description du deuxième scénario-2 : “au cours d'un battement normal du cœur, le sang pauvre en oxygène renvoyé par le corps pénètre dans l'oreillette droite par la veine cave. L'oreillette droite se contracte, ce qui fait passer le sang par la valvule tricuspide ; puis dans les poumons. Simultanément, le sang riche en oxygène provenant des poumons est amené au cœur par les veines pulmonaires. La veine pulmonaire se vide dans la valvule aortique, puis dans l'aorte qui distribue le sang dans les artères de l'ensemble du corps” [10].

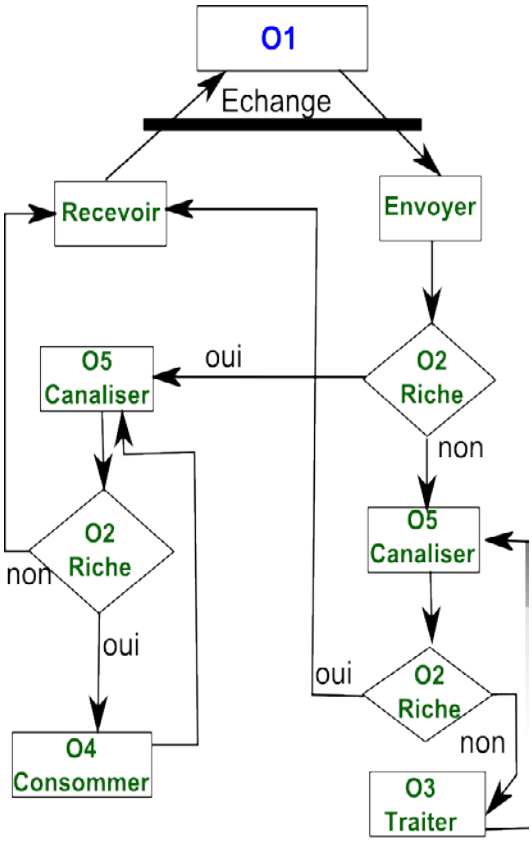

Figure 8 : Logigramme du concept du fonctionnement du cœur humain

À partir de la représentation du scénario-2 (Figure 8), on peut proposer plusieurs projections qui expriment le même scénario, la Figure 9 en est un exemple.

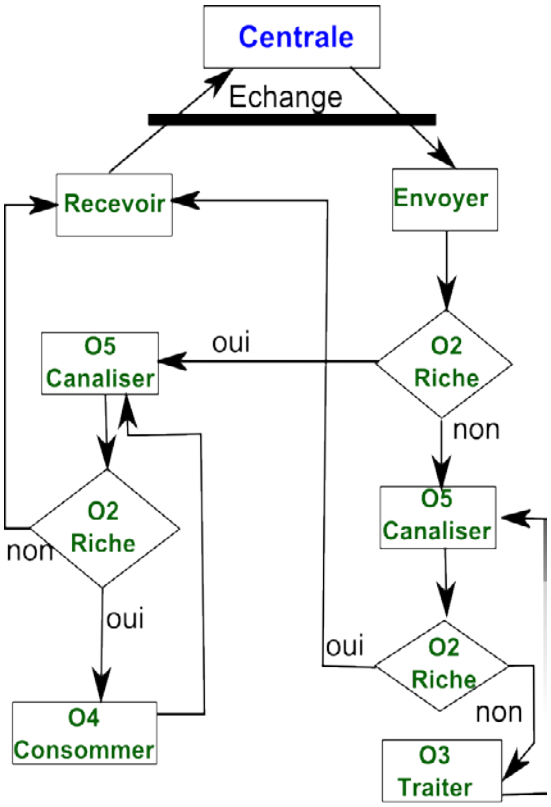

Figure 9 : Logigramme du concept du fonctionnement du cœur humain, dans une centrale de traitement de l'eau.

3.3 Normalisation

D'une projection à une autre, nous avons constaté qu'il y a des éléments qui gardent leur nature et d'autres les perdent. La nature désigne un élément qui est un objet physique ou logique, une action ou un agent.

Quand un élément perd la nature, nous nous intéressons aux concepts et quand on garde la nature de l'élément, nous nous intéressons à la nature de l'élément, de plus de ses concepts.

Un élément peut être décrit par un agent (ex : une personne) ou un objet logique (ex : un compilateur). Dans la figure 6, l'objet de destination a gardé la nature de l'objet source (c'est-à-dire : depuis un agent vers un agent). Par contre, dans l'exemple de la figure 7 l'objet de destination a perdu la nature de l'objet source (c'est-à-dire : depuis un agent vers un

objet logique). Nous pouvons dire que les projections suivent le schéma ci-dessous :

- Éléments : une personne qui conduit une voiture.
- Garder nature : un agent qui contrôle un objet physique.
- Perdre nature : un objet qui contrôle un autre objet.
- Éléments : un développeur logiciel qui produit du code.
- Garder nature : un agent qui produit un objet logique.
- Perdre nature : un objet qui produit un autre objet.

Ces évidences nous ont poussés à trouver une structuration pour les projections, pour une finalité de les exploiter dans un contexte d'apprentissage classique ou numérique. Nous proposons ainsi que, les éléments à enseigner représentent un Niveau de Connaissance du Domaine (NCD), garder la nature représente un Niveau nature des connaissances du domaine (NNCD) et perdre la nature représente un Niveau Concept des Connaissances du Domaine (NCCD).

3.4 Structuration

3.4.1. Niveau de Connaissance du Domaine (NCD)

Dans ce niveau, les connaissances doivent être structurées d'une manière hiérarchique, en spécifiant les objectifs pédagogiques et les capacités cognitives demandées, voir l'exemple dans [7].

3.4.2. Niveau Nature des Connaissances du Domaine (NNCD)

Les connaissances structurées dans le premier niveau doivent être isolées de leur contexte, en enlevant tous les éléments qui caractérisent le contenu du domaine tout en gardant que la nature des éléments. À titre d'exemple :

- Une personne peut conduire une voiture, cela se traduit par agent qui peut contrôler un objet physique.
- Une voiture se compose de moteur, roue, siège, est représentée par un objet physique qui se compose de trois objets physiques.
- Un développeur logiciel produit un code, est représenté par un agent qui produit un objet logique.

3.4.3. Niveau Concept des Connaissances du Domaine (NCCD)

Ce niveau ne garde que les concepts sans faire attention à la nature. L'exemple du deuxième niveau devient :

- Une personne peut conduire une voiture, cela se traduit par un objet qui peut contrôler un objet,
- Une voiture qui se compose de moteur, roue, siège, est représentée par un objet qui se compose de trois objets.
- Un développeur logiciel qui produit un code, est représenté par un objet qui produit un objet.

3.5 Exploitation

Ces trois niveaux vont nous servir comme base pour élaborer des stratégies d'apprentissage et d'évaluation (Figure 10).

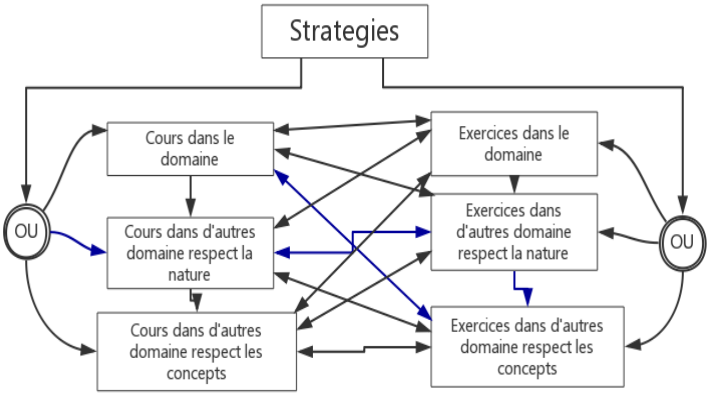
Strategies
Cours dans le domaine
Exercices dans le domaine
OU
Cours dans d'autres domaine respect la nature
Exercices dans d'autres domaine respect la nature
OU
Cours dans d'autres domaine respect les concepts
Exercices dans d'autres domaine respect les concepts

Figure 10 : Les trois niveaux de structuration de connaissances et les stratégies d'apprentissage et d'évaluation.

Une stratégie d'apprentissage combine entre l'apprentissage (soit dans NCD, NNCD et NCCD) et l'évaluation (soit dans NCD, NNCD et NCCD). L'ordre d'exécution peut démarrer par un apprentissage ou par une évaluation, cela dépend du résultat de l'apprenant ou/et de la stratégie souhaitée ou/et de la complexité du domaine à enseigner.

Par exemple, si un apprenant trouve des difficultés de compréhension dans le premier niveau NCD, il faut basculer au deuxième niveau NCCD qui réduit la complexité du domaine à enseigner, car l'information est réduite par rapport au premier niveau. Si cette difficulté persiste, il est possible que l'apprenant à des problèmes avec la compréhension des concepts du domaine, il faut encore basculer vers le troisième niveau NCCD.

Nous croyons que pour chaque apprenant, il existe au moins une stratégie d'apprentissage favorable pour son apprentissage (ex : Figure 10, flèche en bleu), mais pour la prouver, il faut faire plusieurs tests.

La structuration de connaissances proposées dans ce document, peut être utilisé dans plusieurs systèmes, à titre d'exemple, le recrutement, les systèmes d'apprentissage numériques (e-Learning, les Massive Open Online Course 'MOOC') et les Learning Management System, en plus dans l'apprentissage classique.

4. Conclusion

Cet article introduit la nécessité de structurer le contenu à enseigner en trois niveaux, visant à développer des stratégies d'apprentissage facilitant l'apprentissage et des stratégies d'évaluations assurant la compréhension des apprenants.

Le premier niveau définit les connaissances qui ont un lien fort avec le contenu du domaine à enseigner, le deuxième niveau ne garde que la nature des éléments et le troisième niveau présente que les concepts.

Ce travail est le résultat d'une analyse des pratiques des enseignants qu'utilisent l'exemple pour expliquer une connaissance complexe.

La limite de cette approche est que la qualité de l'apprentissage dépend des compétences de l'éducateur ou l'expert qui définit la structure du contenu dans chaque niveau. De plus, le choix des projections doit prendre en considération les aspects culturels des apprenants.

Références

[1] Hattie, J. (2013). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge.

[2] Tardif, J., Meirieu, P. (1996). Stratégie pour favoriser le transfert des connaissances. Vie pédagogique. Vol. 98, N° 7.

[3] Mayer, R. E. (2002). Rote versus meaningful learning. Theory into practice. 41(4):226-232.

[4] Sloutsky, V. M., Kaminski, J. A., Heckler, A. F. (2005). The advantage of simple symbols for learning and transfer. Psychonomic Bulletin & Review. 12(3):508-513.

[5] Paquette, G. (2002). Modélisation des connaissances et des compétences: un langage graphique pour concevoir et apprendre.

[6] Moreno, R., Ozogul, G., & Reisslein, M. (2011). Teaching with concrete and abstract visual representations: Effects on students' problem solving, problem representations, and learning perceptions. Journal of Educational Psychology. 103(1):32.

[7] Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. Theory into practice. 4(4):212-218.

[8] Grosse, C. S., Renkl, A (2004). Learning from worked examples: What happens if errors are included. Instructional design for effective and enjoyable computer-supported learning. pp. 356-364.

[9] Freeman, E., Freeman, E., Sierra, K., Bates, B. (2005). Design patterns: tête la première. O'Reilly.

[10] Webnode (2015). [Online]. Available: http://coeur-articiciel-tpe2013-214.webnode.fr/le-fonctionnement/.

Pedagogical Approaches to Teaching and Learning Entrepreneurship Education

BOUREKKADI Salmane[1], BABOUNIA Aziz[1], SLIMANI Khadija[2], EL KANDILI Mohamed[3], KHOULJI Samira[4]

[1] LRMO national school of commerce and management, Ibn Tofail University, Morocco,

[2] Ibn Tofail University, Kenitra Morocco

[3]Chouaib Doukkali University - FSJES, ELJADIDA Morocco

[4]Laboratory of Information Systems Enginneering Research Group, Abdelmalek Essaadi University, Faculty of Science, Tetouan, Morocco

{salmane.bourekkadi ,ababounia, slimani.uit, m.elkandili }@gmail.com

skhoulji@ac.uae.ma

Abstract

This paper explains what is the pedagogical approach to entrepreneurship education in the world. In the beginning, it explains the different types of models and theories, which have shaped the entrepreneurship structure up to now. Later, the different methods of study that include the traditional lecture-based methods and the innovative methods that are more passive. Then, the former methods are explained in the form of comparison and their pros and cons are discussed. A composite strategy of both these methods can be a practical approach toward effective academic entrepreneurship. After that, the paper explains the future challenges of this course in the Moroccan universities, which have led to the slow development of this field. Risk dilemma, dilution effect, maturity trap, research and publication dilemmas and other challenges are discussed. In the conclusion the solution to the above-mentioned limiting factors is a combination of traditional and non-traditional methods of teaching to make the new entrepreneurs readier for the risky world.

1. Introduction

If we look into the historical perspective of entrepreneurship, we discover that it is deeply related to economics in which the primary goal is growing economically and creating employment. Entrepreneurship in the recent decades strengthened its footing in the education sector. Primarily, it aims to inspire the young entrepreneurs to create more jobs and grow economically. Entrepreneurship in the education sector is mostly termed as "pedagogical entrepreneurship". This term is way different from the conventional core economics and business studies. It is basically a composite of two independent terms, which sets contrast with the previously established practices, and has its own traditions and an outlook on one hand and pedagogy on the other. It has long lasting effect on the socialization learning, motivation, upbringing and formation of more focused education, while entrepreneurship roots itself in the development of the business through initiatives at individual levels and taking risks and their management. In education, entrepreneurship focuses on training people for creating business for themselves. It also emphasizes on a wider concept of educating skills and attitudes for entrepreneurship in order to improve qualities at personal level. It is not mainly focused on creating new business but to honing skills. In 2003, Røe Ødegård tried to explain pedagogical entrepreneurship as a method of teaching, which is more action oriented. In a social context, in this type of teaching, the learner is more actively learning where the personal skills such as knowledge work as foundation providing a direction to the process of learning. This enhances the chances of the development of personal skills in order to analyze better and produce more dynamics by creativity, flexibility, proactivity and cooperativeness. These aspects regulate self-learning implying to setting better goals, planning out the progress and monitoring it at regular intervals so that they can adapt to newer strategies of learning to task stresses. This will yield better and more practical entrepreneurs who can teach the upcoming generations better.

In the previous decade, we saw a commendable evolution worldwide in the research in entrepreneurship. It gained an extensive recognition for being the driving pivot of the economics and society of different nations. The policy makers and scholars today, recognize the value of entrepreneurship education. They range from general to specific and immediately measurable to very complex. It has sprung many long terms as well as short spanned benefits in the social status of the society. These

programs put forward different kinds' objectives for economic growth. When, we will identify the various motives of this type of education, we will have better understanding of the educational requirements. Moreover, we will also be able to make more weighted choice for evaluating criteria and pedagogical techniques. Researchers have concluded that there are distinctive elements of learning entrepreneurship that are teachable and non-teachable. In order to maintain success, educating is to highlight the most practical procedure to tackle the teachable skills and also look for the best conscience of the student needs and the skills that can be taught.

2. Models and theories

There are many theories [1] that are based on different aspects. They are defined as follows.

2.1. Economic Theories

These theories date back to eighteenth century, where Richard Cantillon defined the entrepreneurs as those who take risks. The neoclassical and classical Austrian schools have presented explanations about how entrepreneurship's elemental objective is the main focus at the opportunities and the economic conditions created by them. However, the economic theories are criticized since they fail to recognize the open nature of the markets and the dynamics. They have somehow failed in realizing the uniqueness in the entrepreneur activity and the modulating of the varied contexts in which it occurs.

2.2. Resource based theories

These theories are focused on the leverage at the individual levels from different kinds of source in order to get the entrepreneurial hard work off the ground. They suggest that the easy access to capital enhances the chance for getting-off the ground newer ventures. However, the entrepreneurs start their business with very less capital at hand. Social networking and information are other kinds of resources that the entrepreneurs might leverage. The intangible resource of leadership sometimes yields to a mix operate for the entrepreneur which can't be replaced by business.

3. Psychological theories

This includes the emotional, mental and individual elements to drive an entrepreneurship. A psychologist, namely David McClelland, who is a professor at Harvard puts forward his opinion that the entrepreneurs require achievements to act as driving force for their activities. Julian Rotter another emeritus professor at Harvard has stated a loci of control theory, which says that people with strong belief on their actions can influence people at large, and research has proved that many successful entrepreneurs poses this trait. Another approach suggests that traits of the personal such as creativity, optimism and flexibility drive the entrepreneurial wheel.

4. Anthropological theories

Anthropologists theories are rooted in the explanation of the entrepreneurship at different social backgrounds, which allow the entrepreneurs to leverage different opportunities. A researcher at the Washington university, namely Paul D. Reynolds [2], has pointed out four such social contexts that are ethnic definition, a want for purposeful life, socio-political factors and the social networks. This kind of model looks for the purpose of entrepreneurship by seeing it in cultural perspective and deducing how the forces of culture such as the attitudes of the society has modified the behavior and shape of entrepreneurs and entrepreneurship.

5. Opportunity-Based Theory

Peter Drucker, corporate consultant and a business management author, has stated an opportunity based-theory that postulates the entrepreneurs excel by examining and taking advantage of the possibilities, which are created by cultural, social and technological changes. For instance, if a business is catering the old citizen might see a chance to initiate a new club, if there is a sudden influx of the younger generation as potential death stroke in the neighborhood.

6. Teaching methods

In work of [3; 4; 5; 6; 7], they have classified the methods of teaching as follows:

- Case Study
- Group Discussions
- Group Projects
- Individual written reports
- Action learning
- Individual presentation
- Web Based learning
- Seminars, video recorded learning
- Guest speakers
- Formal lectures and etc.

Another team of researchers, Hytti [8], concluded differently. They believe that the objectives of the entrepreneurship basically decide the medium of education that has to be sought. If the person aims at learning what actually entrepreneurship is, then media, seminars and lectures can be helpful to increase the understanding. These mediums of teaching target a greater audience in a shorter period of time. In case, the primary aim of education is honing the entrepreneurial skills, so the best way to teach is by involving them directly to the processes of entrepreneurship by involving them in training sessions at industries, because these are directly related to work. Another approach is by training of entrepreneurs in controlled environments by role playing or business stimulation.

7. Comparison between the traditional and innovative methods

Authors have categorized the teaching methods as traditional ones, which are more lecture based, and innovative methods, also known as the passive methods, which comprise of mostly action-based strategies. When a comparison is made, we get different point of views of different researchers. According to Bennet, innovative methods are passive and their effectiveness is also limited in having an impression on the entrepreneurial aspects. In 2000, Fiet explained that the instructors also count on the lecture-based methods of teaching, because they are easier to accomplish and require less investment. While, the innovative methods [9] require more time and proper setups in case of controlled environment training. They are believed to nurture better entrepreneurial attributes in the participants. The limiting factor of the traditional methods of teaching are that they make the participants dormant and less encouraged. Although

such students know how to work for entrepreneurs but they hardly have the attributes to become one. In order to make this education completely effective, this teaching should be more of traineeship where the traditional methods are limited to the extent of giving the student commercial reinforcements and on the other hand it should be combined with opportunities to converse, investigate, discuss and question.

8. Challenges in the universities

Today, in Moroccan universities, entrepreneurship education is facing a number of challenges across the globe in the universities. Today, entrepreneurship has become the part of the mainstream. Hardly there is a countable number of trained entrepreneurs, who will be moving forward to setting up business so this might limit the further teaching methods in the coming years. A detailed overview of the problems faced by the entrepreneurship are given as follows.

8.1. The maturity trap

A vast number of universities and business institutes are offering entrepreneurship education, which might lead to a state of saturation in this field. According to the statistics, approximately, all association to advance collegiate schools and about 1000 non-accredited schools are offering majors in the field of entrepreneurship. Although this significant number of institutes wage small battles amongst them. However, there is no proper maturation process in the faculty for this major in the institutes. No proper departments are given for this fields in the universities. The faculty and entrepreneurship journals are not there and recognized widely. That is why a thorough academic legitimation of this area is yet unexperienced and this field still needs to mature.

8.2. Research and publication dilemma

Katz [1] has highlighted two problems in the entrepreneurship over abundance. Too much literature yielding only few valuable papers and regular pushing for researchers of great potential to publish in the mainstream journals of the management. Both of these aspects may be seen as opportunities as well. An increase in the number of publishing papers in the important journals will proportionally increase the number of scholars in the review boards in the field of entrepreneurship. If the

entrepreneurship faculty will push the ranking to limits of the mainstream journals, we will be able to produce more quality research venues to generate literature. The simple accepted fact is that a business school is run by the research it has conducted. Thus, research should be an elemental portion of entrepreneurship, which is accepted as well as respected.

8.3. Faculty shortage

This problem is also deal at dual levels. The first problem being faced is that there is a shortfall of teaching faculty at academic levels, and secondly there is a lack of PHD in philosophy program to provide basic entrepreneurship assistance. For this, more equipped business institutes should be set up to cover the gap of PHDs in the field of entrepreneurship. Leading schools such as Colorado University, University of Georgia and Case Western Reserve University should begin programs. Until these programs are developed the faculties might be trained accordingly as an extra effort. A lack of faculty at various levels can be overcome by improving the acceptability and respect of the entrepreneurship journals to promote the faculty. This is how facultative staff at many ranks will increase

8.4. Technology challenge

The teaching methods should evolve according to the advancement in the technology, by using them into the process, because this will yield better results, which will transform the education system.

8.5. Business vs academic incongruence

Students are limited to classrooms to story making mostly although they should be interacting with the entrepreneurs, who have had faced the challenges already. They should delve themselves in the issues and real problems to make a difference

8.6. Risk dilemma

Many risk factors are involved such as financial, family, career, social or psychic risks. Certainly, these risk factors are important and they can't be neglected. Although the entrepreneur is a calculated risk taker and he deliberately delves himself in the moderate risks instead of behaving as a

high-risk gambler of mythical type. Worse situation is when we lack educators, who are not ready to risk all in their curricula or programs of entrepreneurship. Many faculties have focused on tenure and they leave the challenges for later career life. Younger faculty is required to pursue their career in the academic entrepreneurship.

8.7. Dilution effect

In many universities, entrepreneurship has been vastly legitimized to such an extent that the researchers fear this might dilute its real meaning. Many disciplines are associated with word as a prefix. However, they want to make sure that those disciplines should be related to the tittle instead of using the name merely.

9. Conclusion

Entrepreneurship is not only limited to risking. Instead, it creates awareness for the initiation and survival of business in different competitive environments. The teaching methods mentioned above either conventional or innovative need to be applied in combination because either of them has its own pros and cons. The risk factors described above are although hindering the growth of the pedagogical entrepreneurship education in the universities. However, they can be overcome by improving the faculties and the technological availability. The solution to all the problems is that for improvement in the teaching methods. The teaching methods should be reviewed and enlisted. This should be further researched qualitatively by the researchers to identify the key teaching methods and complete the list. It has been concluded through the comparisons that problem solving, case study, and further action-based teaching methods are most appropriate in the course of learning entrepreneurship. Accustoming the students with the real challenges of the business world, and training them in mock business environments can improve the academic structure further.

It is very important to note that entrepreneurship methods and theories require practice. In our minds learning, the model and theories are more important that anything. However, in this ever-changing world the challenges are vast. Hence, such methods should be taught, which can stand out the trials of drastic fluctuations in the context and content. It would be appropriate, if we say that while teaching, we don't teach

entrepreneurship rather we teach a method in order to navigate and explore this discipline.

Acknowledgment

The authors wish to thank Souhail BOUREKKADI, Youssef EDDABARH, Bouthaina EDDABARH. This work was supported in part by a grant from ERISI.

References

[1] Katz, J. (1992). Modeling entrepreneurial career progressions: Concepts and considerations. Entrepreneurship Theory and Practice. 19(2):23-39.

[2] Reynolds, P. (1992). Sociology and entrepreneurship: Concepts and contributions. Entrepreneurship Theory and Practice. Vol. 16, pp. 47-70.

[3] Brockbank, A., McGill, I. (2007). Facilitating Reflective Learning in Higher Education, 2nd ed. New York: Open University Press.

[4] Plaschka, G. R., Welsch, H. P. (1990). Emerging Structures in Entrepreneurship Education: Curricular Designs and Strategies, Entrepreneurship Theory and Practice. 14(3): 55–71.

[5] Venkataraman, S. (1997). The Distinctive Domain of Entrepreneurship Research, in Advances in Entrepreneurship, Firm Emergence and Growth. Eds. G. T. Lumpkin and J. Katz. Greenwich, CT: JAI Press. pp. 119-138.

[6] Arasti, Z., Falavarjani, M. K., Imanipour, N. (2012). A Study of Teaching Methods in Entrepreneurship Education for Graduate Students. Higher Education Studies. 2(1):2-10.

[7] Arasti, Z., Kiani Falavarjani, M., Imanipour, N. (2011). Teaching methods in entrepreneurship education: the case of business students in Iran. In Proceedings of the 6th European Conference on Innovation and Entrepreneurship. Vol. 1, pp. 92-98.

[8] Hytti, U. (2008). Enterprise education in different cultural settings and at different school levels. The dynamics between entrepreneurship, environment and education. Vol. 131.

[9] Ahmadi, A., Mohammad kazemi, R., Elyasi, G. M. (2017). Identification of entrepreneurship teaching methods in education's affective domain through edutainment approach. Journal of Entrepreneurship Development. 10(2):11-27.

A Dialogue-System Using a Quranic Ontology

BENDJAMAA Fairouz[1], TALEB Nora[1]

[1] Computer science Department, University of Badji Mokhtar, Annaba, Algeria

{Bendjamaa, taleb}@yahoo.com

Abstract

A dialogue-system is a medium of communication and interaction between human and computers by exchanging questions and answers in natural languages. The performance of the latter depends on the ability to analyze the question; the good management, which allows choosing the keywords to question the knowledge base; and finally the use of an efficient knowledge base. However, several problems related to developing dialogue systems are due to the use of traditional databases. In this project, we propose a dialogue-system based on Quranic ontology. This system allows easy access to Quranic information by SPARQL queries from pre-existing domain ontology. The used ontology covers Quranic chapters and verses, each word of the Qur'an and its root and lemma.

1. Introduction

A dialogue-system is a human-machine interaction in natural language in order to start a search in a given domain. To accomplish a task of searching for specialized information by interacting with the user in natural language, the machine must try to analyze what the user is looking for and propose answers. Such a system examines with the user the keywords that might be used. It hence tests several queries using different criteria to orient the search in the direction desired by the user.

The structure of a simple knowledge base has troubles managing all the dialogue-system requirements. As a result, some developer tried to integrate the notion of ontologies as the basis of the dialogue-systems.

After the appearance of the semantic web and ontologies many dialog-systems based on ontologies have been created. The structures of ontology explicitly represent a domain knowledge using a machine-understandable format and they can be incorporated into computer applications and systems, facilitating annotation data, decision making, information retrieval and natural language processing.

The problem of understanding the natural language at the conceptual and practical levels is still in the first stage. There are, for example, annual competitions such as the Loebner Prize or the Chatterbox Challenge aimed at passing a test and thus imitating human verbal interaction, but no program has succeeded in reaching the level of a human [1].

This paper is laid as follow: Section 2 gives a quick state of the art of the related notions and similar projects. Section 3, briefly, presents our project. Section 4 details more our implementation and the actual results and state of progress and we finally conclude this paper at Section 4.

2. State of the art

2.1. What is a dialogue-systems?

The idea of replacing the human interlocutor with a machine has emerged since the 1960s, when many computer systems were designed with this objective in mind. Even if communication with others proves to be a complex mechanism whatever the situation and the context, a dialogue with a machine is a subject of interest [2].

In its first definition, dialogue is a "conversation between two or more persons on a defined subject". By adding the term "dialogue system" is the idea that at least one of the participants in this conversation is a Machine in the broad sense of the term. In this sense, the Glass ‘work [3] give the following definition: The term dialogue-system generally indicates a system allowing interaction between a human and a system within a restricted framework. Nevertheless, since this interaction can take many

forms (textual chat, voice server, virtual agent, etc.), it is difficult to draw up an exhaustive state of the art [4].

A dialogue-system can be declined in many forms with varied characteristics. This explains why it is difficult to find a unified architectural standard in the literature. Many proposals exist, for example, open-domain approaches such as those presented in [5] and [6]. The latter are mainly based on the exploitation of general knowledge sources such as the semantic web to allow the system to cover a large thematic field.

2.2. What are the components of a dialogue-system?

Different dialogue systems have different architectures, but they have the same set of components which are [7]:

- **Input decoder:** It is the component responsible of recognizing the input by converting it to a String. This component is only present in non-text-based dialog systems. It involves converting the spoken sound (user utterances) into text (a string of words). This requires knowledge of phonetics and phonology. In addition to speech, the dialogue system may have other inputs such as gesture, writing, and so on.

- **Language understanding:** As it says, this component tries to understand what the user means. It converts the user's query into a semantic representation using morphology, syntax, and semantic.

- **Dialogue manager:** The dialog manager supervises the dialog. It treats the semantic representation created by the language understanding component, determines the context, and creates a semantic representation of the response. It performs many tasks such: keeping the history of the dialogue, adopting some dialogue strategies, treating malformed and unrecognized text, recovering the content stored in files or a database, deciding on the best answer for the user, … etc. The dialogue manager has different models to deal with these varied tasks, like the dialogue model, the user model, the knowledge Base, the discourse manager, the reference resolver, and the grounding Module.

- **Response generator:** It constructs the message containing the answer for the user's query. It decides which information to use and

how to structure it to create the message. Current systems use simple methods like the insertion the extracted data into a predefined template.

- **Domain specific component:** This component is responsible for converting the user's query from the internal representation used by the dialogue manger to the representation used by the external specific system and hence allowing the dialogue manager to interact with external software. It generates SQL queries from natural language text.

- **Speech generation:** It translates the message constructed by the answer generation component into spoken form. For speech generation.

2.3. What is an Ontology?

Ontology models the knowledge of a specific field, and this modeling consists of concepts, the relationships between these concepts, axioms and instances.

A concept is the description of an action, a strategy or reasoning process. The relationships define the type of interaction between concepts [8]. The instances are the individuals of the domain represented by the ontology and the axioms are used to model assertions or statements that are always true [9].

2.4. Languages and tools for ontology construction

Several languages and tools have been developed to represent and manipulate the ontologies, among these languages: DAML+OIL (DARPA Agent Markup Language+ Ontology Inference Layer) and the standards that emerged from the World Wide Web Consortium W3C such as: XML (Extensible Markup Language), RDF (Resource Description Framework), RDFS (RDF Schema) and OWL (Web Ontology Language).

- DAML+OIL as indicated by its name, it has been developed by the joint from the US project DAML, which is a semantic language of marking for the resources of the Web, and the European Union project OIL that extends RDFS by offering new primitives for the definition of the classes [9].

- XML, a recommendation of the W3C, it was developed, in 1996, to describe and exchange data on the web [10].

- RDF is based on XML. It describes knowledge as web resources to facilitate their automatic processing. RDFS is a semantic extension of RDF allowing the representation of resources and the relationships between these resources [10].

- OWL was built by W3C, in 2001, inspired by DAML + OIL. It offers a rich vocabulary and formal semantics for the description of ontologies. [9].

Also, many tools based on these languages are used to implement ontologies, we enumerate: Protégé and Ontolingua.

- Protégé is a free, open-source, and a very popular tool. It was developed by the Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine. Protégé allows editing, visualization, control, and fusion of ontologies. [11]

- Ontolingua is an ontology server created at the Knowledge System Laboratory at Stanford University. It allows a user or group of users to view, create, merge and reuse ontologies with the ability to export them in multiple formats [12].

2.5. Related works

A Method for designing Dialogue Systems by using Ontologies, presented in [13], is a methodology for designing dialogue-systems, turning easier the task of building the knowledge base for a dialogue system. A method to design and aggregate existing ontologies in these systems are also being proposed. For this reason, pattern matching, natural language processing tools, thesaurus and the language AIML were also used.

Generic natural language command interpretation in ontology-based dialogue systems, presented in [14], is a general architecture towards a more generic approach to build conversational agents. This architecture contains generic (in sense of application independent) natural language (NL) modules that are based on ontologies for command interpretation. It focuses on the presentation of the event generator and dialogue manager

modules, which rely on a bottom-up approach for matching the user's command with the set of currently possible actions.

The Arabic chatbot giving answers from Qur'an, presented in [15], uses machine learning techniques to produce an Arabic chatbot, which accepts input in Arabic and produces responses from the Qur'an. A system that learns conversational models of a transcribed corpus of conversation has been used to produce a range of chatbots that speak various languages, including English, French and African. In principle, Since the Qur'an is not a transcript of a conversation, they adapted the study process so that the chatbot could handle the structure of the Qur'an with regard to the chapters and verses.

The architecture of this system divides the program into three parts. The first part creates the Qur'an frequency, the second part generates the original template and template files, and the third part implements the restructuring phase and generates the AIML file.

The dialogue-based Visualization for Quranic Text "AQILAH", presented in [16], uses an architecture that represents the dependencies and inevitable relationships between the different components of the knowledge base in the system, which facilitates the user's understanding of the internal interaction in the visualization system based on dialogue. AQILAH is equipped with a minimal parsing technology, based on a single keyword to analyze the textual input of human natural language. It responds in a natural language using text from Qur'an by extracting the keywords from an input string and searching for the keywords in the knowledge base, a dictionary of associative words and a backup of random answers.

A Simplified Approach Towards the Development of Natural Dialogue Systems "Nadia", presented in [17], is a set of components that deals with the creation of spoken dialogue systems. While common standards like Voice XML are widely used in the industry, it is still difficult to design dialogues that use well-established theories from research projects. This software (created in 2013) is part of a PhD thesis. Its main purpose is to facilitate the creation of dialogue systems and to show the effect of different dialogue strategies, which is not only important in industry but also in university courses.

3. Our project

Our system is loaded with several tasks that seem quite complicated, since it is an interaction in Arabic natural language. From a distance, it is obvious that we must first analyze the user's input, passing by the segmentation, and then creating a semantic path. In addition, we also need a module to access the ontology to extract the information.

To reach these results, we propose the following approach:

- The system accepts as input a spelling string in Arabic (entered by the user).
- The orthographic chain is then received by a linguistic component (called comprehension component) that provides a semantic schema representing the literal meaning of the utterance.
- The result of the linguistic component, then, enters in an iterative process between the dialogue manager and the query engine that can access our ontology, in order to refine the query and obtain a precise answer that satisfies the user.
- The result obtained from the ontology must be generated in natural language.
- Display the result for the user in natural language.

3.1. Web system

Before modeling our dialogue-system, we proposed a general architecture of a general web system that can be applied to any dialog-system. This architecture was inspired by the model [13].

In general, we get an architecture composed of an application (dialogue-system) in a server with a web interface. In the following, we will be interested in the architecture of the dialogue-system.

3.2. Dialogue-system

To accomplish the task of searching specific information by interacting with the user in natural language, the machine must attempt to analyze its

objective and offer a suitable response by offering some examples or assistance to lead the interlocutor towards a solution. Such a system examines with the user the keywords that might be used to interrogate the knowledge base. It thus tests several queries using criteria to orient the search in the direction desired by the user. In combination with domain ontology, the path to the solution will be shorter and more efficient.

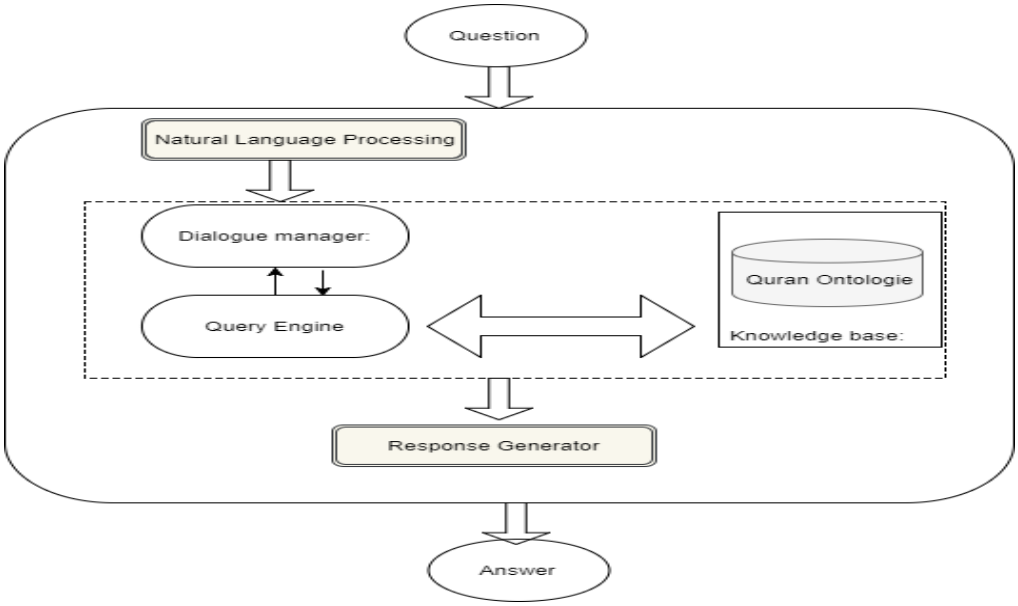

Figure 1: The architecture of our dialog system based on an ontology of quran

- **Natural language processing:** The recognition module sends a sequence of words to the literal interpretation module that performs an out-of-context analysis of this word sequence and produces a semantic scheme.
- **Query engine:** This module is the bridge to the knowledge base. It contains queries on the different concepts and relationships of the ontology to extract the correct answers.
- **Dialogue manager:** This module is responsible for choosing the words on which the system must rely to question the knowledge base. Thus, it takes care of the cases of error and work on the misunderstandings in cooperation with the query engine.

- **Response generator:** This module comes last. After the extraction of the response with a given type (everything depends on the ontology). This module takes a semantic structure as input and provides a structure of the same type output but reformulates the result in natural language.

- **Knowledge base:** In our case, it is a pre-existing Quranic ontology. The ontology is a conceptualization of the Qur'an, which conforms to the needs of our system.

3.3. Qur'an ontology

Aimad Hakkoum and Said Raghay have adopted an approach that allows human beings and computers understand the Quranic knowledge all along the creation of a Quranic ontology. The ontology target is to build a computational model capable of representation all the concepts mentioned in the Qur'an and the relations between them using protege owl. The ontology could be questioned with SPARQL requests.

They followed the methodology discussed in [18] by adopting an iterative approach to ontology development with the six steps: define the ontology domain and scope, review existing ontologies, enumerate important terms in the ontology, define the classes and the class hierarchy, define the properties of classes and their facets and finally create instances.

The ontology covers the following subjects: Quranic chapters and verses, each word of the Qur'an and its root and lemma to facilitate the keyword search. It does not cover words morphology search but we will add links to QVOC ontology where this is covered. However, it covers the pronouns in order to define their antecedents. The following figure illustrates the conceptual graph of this ontology.

4. Implementation

In the implementation, we considered the input queries as objects.

- **Client:** the client is the web interface (access with the browser). When running, the question is sent to the server as an HTTP Query.

- **Server:** The server part receives the HTTP Query and sends a question of type "character string" to the application for processing.

In the opposite direction, it receives the response provided by the module 'Application' in JSON format it extracts the requested information and sends the response to the client in a HTTP format. Each time (s)he receives a question or answer during a run, (s)he stores them in an external file dedicated to the dialogue history.

- **Application:** All the treatments are made at this level. The result of this module is in the JSON format. In this module, the semantic path from the input to the output is as follows:

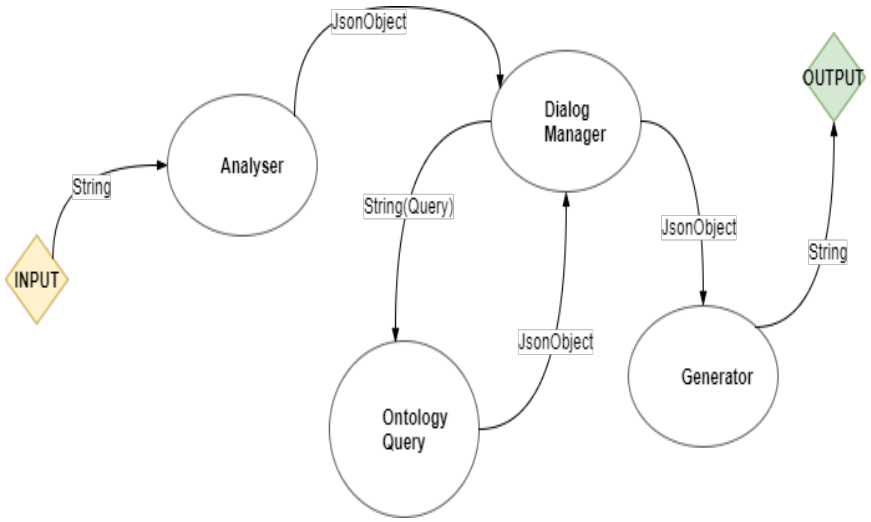

Figure 2: Functional structure of the dialog system (the implementation steps from the Question INPUT to Answer Generation)

- **Analyze:** For the syntactic and semantic analysis of the question, we used the API corenlp Stanford. The result is then transformed into Json format. It passes through the manager for the choice of the key words (String Query) to interrogate the ontology. This part is coded in JAVA.

- **Ontology query:** It is the query engine module; it represents in the implementation the apache Jena API. The queries are written in SPARQL language. The queries were organized as a method for a simple use and reuse.

4.1. Result obtained

By querying the system, it extracts all the verses that have information related to the user's query. Finally, we obtain the following result:

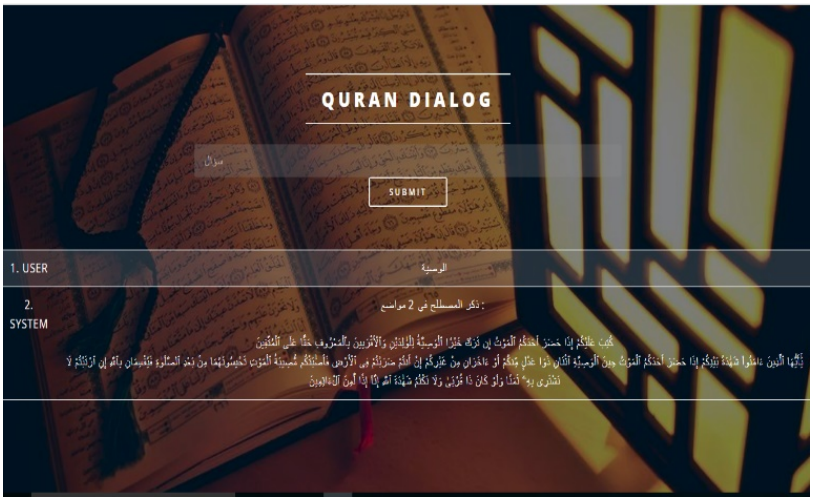

Figure 3: A result when querying the system

5. Conclusion

In this project, we developed a dialogue-system based on an existing Quranic ontology. The ontology used represents a conceptualization of Qur'an, which facilitates the access to the information contained in this sacred book. The realization of this project was done using a semantic analyzer of the Arabic language, a web server, and a query engine. It aims at facilitating the search of Islamic information in an interactive way.

Due to lack of time, as this project was realized as a sub project within a master thesis, we had only two to three months to work on it. Therefore, we were not able to fully complete the project and hence we are planning to further work on it to have a better and more precise answers and further enhance the dialogue manager by upgrading its different models.

References

[1] Floridi, L., Taddeo, M., Turilli, M. (2009). Turing's Imitation Game: Still an Impossible Challenge for All Machines and Some Judges, An Evaluation of the 2008 Loebner Contes, Minds and Machines.

[2] Coulon, D., Kayser, D. (1986). Informatique et langage naturel : Présentation générale des méthodes d'interprétation des textes écrits, Revue TSI : Technique et Science Informatique. Vol. 5. N° 2.

[3] Glass, J. (2000). Data collection and performance evaluation of spoken dialogue systems: the MIT experience, Actes de ICSLP'00, Pekin, Chine.

[4] Ferreira, E. (2013). Apprentissage automatique en ligne pour un dialogue homme-machine situé. Thèse de doctorat, Université d'Avignon et des Pays de Vaucluse, Avignon, France.

[5] Strzalkowski T., Harabagiu. S. (2006). Advances in open domain question answering. Vol. 32. Springer Science & Business Media.

[6] Lcock, G., Wikitalk W. (2012). A spoken wikipedia-based open-domain knowledge access system, Proceeding ICCL.

[7] Arora, S., Batra, K., Singh, S. (2013). Dialogue system: A brief review. arXiv preprintarXiv:1306.4134.

[8] Abderrahim, M. A. (2016). Exploitation des Ontologies dans les Systèmes de recherche d'informations Arabes. Thèse de doctorat, Université Aboubakr Belkaïd, Tlemcen, Algérie.

[9] Barkat, O. (2017). Utilisation conjointe des ontologies et du contexte pour la conception des systèmes de stockage de données. Thèse de doctorat, Ecole Nationale Supérieure de Mécanique et d'Aérotechnique, Poitier, France.

[10] Slimani, T. (2015). Ontology Development: A Comparing Study on Tools, Languages and Formalisms, Indian Journal of Science and Technology. Vol. 8.

[11] Wantroba, E.J., Romero, R.A.F. (2014). A Method for designing Dialogue Systems by using Ontologies, Standardized Knowledge Representation and Ontologies for Robotics and Automation. Chigago, USA. pp. 30-35.

[12] Mazuel, L., Sabouret, N. (2006). Generic natural language command interpretation in ontology-based dialogue systems, Laboratoire d'informatique de Paris 6 - LIP6, France.

[13] Bayan, A.S, Eric, A. (2004). An Arabic chatbot giving answers from quran, JEP-TALN 2004, Arabic Language Processing, Fez, Morocco.

[14] Mustapha, A. (2009). Dialogue-based visualization for Qu'ranic text, European Journal of Scientific Research.

[15] Markus, M. (2015). NADIA: A Simplified Approach towards the Development of Natural Dialogue Systems, Natural Language Processing and Information Systems. pp. 144-150.

[16] Hakkoum, A., Raghay, S. (2015). Ontological approach for semantic modeling and querying the qur'an. In Proceedings of the International Conference on Islamic Applications in Computer Science and Technology.

[17] Protégé Website (2018). [Online]. Available: https://protege.stanford.edu/.

[18] Ontoligua Website (2018). [Online]. Available: http://ontolingua.stanford.edu/.

A Brief Survey on Named Entity Recognition in Amazigh Language

TALHA Meryem[1], BOULAKNADEL Siham[2], HAMMOUCH Ahmed[1]

[1] Laboratory of Research on Computer Science and Telecommunications LRIT, Faculty of sciences Rabat, University Mohammed V, Morocco

[2] Royal Institute of Amazigh Culture, Allal El Fassi Avenue, Madinat Al Irfane, Rabat-Institutes, Morocco

{meryemtalha,hammouch}@yahoo.fr
boulaknadel@ircam.ma

Abstract

Named Entity Recognition (NER) is a core subtask of Information Extraction (IE). The main idea is to extract named entities from a given text and classify them into a set of named entity classes such as person names, locations, organizations, etc. It is an essential part in many applications, such as Machine Translation (MT) and Question Answering System (QA). Amazigh NER has gained considerable attention in the past few years. However, the peculiarities of Amazigh arise the challenges of this task. In this paper, a survey regarding the recent progress made in Amazigh NER research using different techniques is presented.

1. Introduction

Identifying and classifying entities of a given text into different predefined classes, such as names of persons, organizations, locations, expression of times, quantities, monetary values, temporal expressions, percentages, etc., is a process called named entity recognition (NER) [1]. The term Named Entity was first introduced in the sixth Message Understanding Conference (MUC-6). This term covers not only proper names but also includes

temporal expressions, numerical expressions and other types of units depending on domain of interest.

In MUC-6, Named Entities (NEs) were classified into three types of category as follows [2]:

- ENAMEX: person, organization, location ;
- TIMEX: date, time ;
- NUMEX: money, percentage, quantity.

NER also provides basic inputs for various NLP tasks, including:

- Information Extraction ;
- Questions Answering ;
- Automatic Summarization ;
- Machine Translation.

A much of NER work has been done in English and some other foreign languages similar to Spanish, French, Chinese, Arabic, etc., with great precision but NER in Amazigh language is at a primary stage.

This paper investigates the progress in Amazigh NER research. To our best knowledge, Amazigh NER and categorization have not yet been surveyed, which has motivated us to conduct this survey.

The remaining sections of this survey are organized as follows: Section 2 provides background information about NER approaches. Section 3 describes the specificities of the Moroccan Amazigh language. Section 4 discusses the linguistic issues and challenges impeding the extraction of named entities in Amazigh language. Section 5 reports the Amazigh linguistic resources that have been used for the Amazigh NER task. A state-of-the-art in Amazigh NER research is presented in Section 6, and the survey is concluded in Section 7.

2. NER approaches

NER systems have been developed using mainly three approaches: the rule-based approach, the machine-learning based approach and the hybrid approach.

2.1. Rule-Based approach

The rule-based approaches typically make use of two types of resources [3]:

- Handcrafted rules set manually written by linguists;
- A set of gazetteers including the lexical trigger words.

The main advantage of the rule-based NER approaches is that they are able to detect complex entities due to the solid linguistic knowledge. However, the main disadvantage is the non-ability of portability. These approaches give better results on restricted domains and in order to apply them on new ones, it requires new adapted grammatical knowledge and background of the particular domain. However, for low-resourced languages, handcrafted rules remain the preferred technique.

2.2. Machine-Learning approach

Much of the current researches in NER on the well-known language involve machine-learning based approaches. These approaches treat NER as a classification process and make use of a large amount of NE annotated training data [4].

There are two types of machine learning models that are used for NER called Supervised (SL) and Unsupervised (UL) machine learning model. The main difference between these types is that the first one requires the availability of large annotated data in the training stage, while the second one does not need an annotated data beforehand. It relies on clustering similar documents or entities together.

Several ML techniques have been widely used for the NER task of which Hidden Markov Model [5], Maximum Entropy [6], Conditional Random Field [7], Support Vector Machine [8] are most common.

Unlike the rule-based approaches, Machine-learning based approaches can be easily applied to different domain or languages. However, creating large enough training sets for them remains a problem.

2.3. Hybrid approach

The hybrid approach is the combination of rule-based and machine learning-based approaches. The main idea is to make use of the strongest points from each approach and optimize the overall performance [2].

3. Amazigh Language

Amazigh language belongs to the branch of the large Afro-Asiatic (Hamito-Semitic) linguistic family [9-10]. It covers a boundless geographical zone: all of North Africa, the Sahara (Tuareg), and a part of the Egyptian oasis of Siwa. In Morocco, Amazigh language is one of the national and official languages besides the classical Arabic. According to the last census of 2014[1], it is spoken by close to 27% of the population. Moreover, three different dialect clusters have been differentiated: Tarifite (4.1%) in North (Rif), Tamazight (7.6%) in Central Morocco (the Mid-Atlas and a part of the High-Atlas) and South-East, and Tachelhite (15%) in the South-West and the High Atlas.

4. Linguistic issues and challenges of Amazigh language

Amazigh is a highly concatenative language and it makes Amazigh language morphologically rich with very productive inflectional and derivational processes, and it differs from English or other Indo-European languages. Despite the achievements made in Amazigh NER research, the task still remains challenging due to many issues.

For example, this language does not have capitalization, which is a major feature used by NER systems for European languages. Another issue is that the official alphabet for Amazigh language in Morocco is “Tifinaghe”, which is different from Latin Alphabets (ⴰ [a]; ⴱ [b]; ⴲ [c]; ⴷ [d]). However different scripts have been frequently used to write the Amazigh language such as Latin and Arabic scripts.

Another issue is the lack of available Semantic and Linguistic Resources for Amazigh NER, corpora and lexical resources are the two main types of linguistic resources. Among other problems, one last example is the lack of standardization and spelling variation. The Amazigh text does not respect the standard writing convention. For example, the person name ("ⵜⵉⵎⵣⵉⵖⵉ,

[1]http://www.hcp.ma/Presentation-des-premiers-resultats-du-RGPH-014_a1605.html

bnkiran, Benkiran") can be written as ("ⵏⵏ ⴽⵉⵔⴰⵏ, Bn Kiran, Ben Kiran"), and the Location name ("ⴼⵇⵉⵀ ⵏⵏ ⵙⴰⵍⵃ, Fkih Ben Saleh") that can be-written as ("ⴼⵇⵉⵀⵏⵏⵙⴰⵍⵃ, fkihbnsaleh").

5. Amazigh Linguistic Resources

Amazigh is a low-resource language, however with the growing interest in Amazigh NER research, some efforts has been made in creating standardized linguistic resources in order to facilitate the development of Amazigh NER systems. This section discusses the various resources created.

5.1. Corpora

As mentioned before, the main problem of Amazigh language is the lack of publicly available annotated corpora. Therefore, some researches have built their own corpora for training and testing purposes. The last updated version of the Amazigh corpus "AMCorp" contains more than 900 news articles published online[2] that are collected from a broad range of topics (sports, economics, news on royal activities of His Majesty King Mohammed VI, and many others), containing news that happened over a period of 2 years (dated between May 2013 and July 2015). The articles are selected in such a way that the data set contains different types of information, and that the system's future use will not be limited to any particular text type. It consists of nearly 170.000 words, after some data cleaning operations like deleting non-Amazigh words. This data set is manually annotated following the MUC guidelines, ENAMEX (Location, Person, and Organization), NUMEX (Numbers, Percentage and Money) and TIMEX (dates & times) types.

5.2. Gazetteers

As far as the corpus is concerned, the gazetteers are also needed in Amazigh NER systems. They are considered one of the important Amazigh linguistic resources, a description of the Amazigh gazetteers is given below:

[2] http://www.mapamazighe.ma

- Person gazetteer: consists of famous person names splited into first name and last name gazetteers in order to identify different combinations. It contains 2533 entries.

- Location gazetteer: includes different location names in Morocco, with names of almost all the countries in the world, cities, states, and geo-graphical names from different sources. The total of entries is 2318.

- Organization gazetteer: contains names of important organizations such as those of political parties, universities, agencies and banks. 913 entries have been collected.

- Date/Time gazetteer: includes entities related the temporal expressions, such as days and months entities. The majority of these entities were extracted from the IRCAM[3] website. This contains 193 entries.

- Additional gazetteers of numerical expressions including numbers (216 entries), money (14 entries) and percentage (3 entries) have been created.

- Lists of 468 trigger words have also been created manually, which generally provide cues surrounding the named entities that would indicate their presence, it contains titles like (ⵎⵓⵙ, Mass, Mr) and (ⵜⴰⴱⴰⵔⵜ ⵏ ⵜⵜⴰⵣⵔⵜ ⵏⵡⴰⵍⵍⴰⵏ ⵓⵡⵍⴰⵏ ⵎⵓⵍⴰⵢ, bab n tattuyt tagldant agldun mulay, His Majesty the King).

6. Amazigh NER Systems

In this section, we present different Amazigh NER systems. They are classified according to the approaches used. But before, we should mention that all experiments have been done using GATE.

[3] www.ircam.ma

6.1. GATE

A good number of tools are available for developing and evaluating NER systems. We have chosen GATE[4] because it's one of the most popular freely available software tools dealing with NLP. The tool supports nine languages (English, French, German, Italian, Chinese, Arabic, Romanian, Hindi, and Cebuano) [11-12]. It provides a framework, which the development of the rule-based NER systems is easy. The user has the ability of implementing grammatical rules as a finite state transducer using JAPE, even the machine-learning based systems can be implemented using GATE.

6.2. Rule-Based systems

One of the first research papers in the field was presented by Talha *et al.* [13]. The paper describes an Amazigh rule-based NER system. It is able to extract and recognize person names, locations, organizations. It relies on a set of 17 grammar rules and 3710 lexical resources. For evaluation, 200 texts from AmCorp were selected randomly and manually tagged. The overall performance obtained for the following categories: person, location, and organization was respectively as follow: F-measure of 64%, 40% and 82%.

As a continuation of the initial attempt, Boulaknadel *et al.* [14] developed an enhanced rule-based Amazigh NER system. The system identifies the following NE types: person names, locations, organizations, date and numbers. For the experiment, the authors used around 289 news, the size of gazetteers was 4666 entries. They reported an f-measure of 83% for person names, 97% for locations, 76% for organizations, 67% for dates and 95% for numbers.

Another NER system adopting the rule-based approach for recognition and classification of Amazigh named entities is presented by Talha *et al.* [15]. In this research, the authors added some gazetteers (5193) and grammar rules (76) to the system to increase the performance. They applied the set of rules and gazetteers on a corpus containing 430 news. The system was able to recognize five classes of named entities. It obtained an F-measure of 81.5% for person, 87.75% for location, 84% for organization, 80% for date, and 83.5% for numbers.

[4] General Architecture for Text Engineering, https://gate.ac.uk/

6.3. Machine-learning systems

The work of Talha *et al.* [16] is a new attempt to improve performances of the previous Amazigh NER systems.

The authors tried to recognize the Amazigh named entities using a supervised machine learning approach (using SVM [17]) and exploring different sets of features. The features include token form, token kind, semantic classes from gazetteer lists and named entity types.

The system is able to identify the following NE types: person, location, organization, numbers, date/time, money and percentage. The overall system performance in terms of F-measure was as follows: 81%, 82%, 86%, 88%, 94%, 94 and 100%, respectively using training set of 800 texts and test set of 100 texts.

6.4. Hybrid systems

Recently, Talha *et al.* [18] proposed a hybrid NER system for Amazigh. The rule-based component is a duplication of the NERAM system [15]. The ML-based component uses the SVM classifier. The feature set used includes the NE tags predicted by the rule-based component, contextual features and the gazetteers features.

The system identifies the following types of NEs: person names, locations, organizations, dates and numerical expressions. The overall performance obtained for the various categories using AmCorp was an f-measure of 73%. The experimental results showed that the hybrid Amazigh NER approach didn't attempt a very good improvement of results compared to the rule-based and the ML-based components when they are processed individually. This is due to the minimized feature set used, the lack of POS tagging and morphological features.

7. Conclusion

This paper has presented a brief literature review of the major works done regarding the concept of named entity recognition for Amazigh language. Amazigh NER works are in progress, the number of current Amazigh researches is still insufficient compared with the others well-resourced languages due to many issues such as the lack of a huge annotated corpora, lack of capitalization, variations in writing style and difficult

morphology. Our main aim is to provide a key to deal in some detail with Amazigh NER research and guides researchers in interesting and fruitful research directions.

References

[1] Nadeau, D. (2007). Semi-Supervised Named Entity Recognition: Learning to Recognize 100 Entity Types with Little Supervision, Ottawa Carleton Institute for Computer Science, School of Information Technology and Engineering, University of Ottawa.

[2] Grishman, R., Sundheim, B. (1996). Message Understanding Conference - 6: A Brief History. In Proc. International Conference on Computational Linguistics.

[3] Nadeau, D., Sekine, S. (2007). A survey of named entity recognition and classification. Lingvisticae Investigationes. Vol. 30, N°1. pp. 3-26.

[4] Sharnagat, R. (2014). Named Entity Recognition: A Literature Survey.

[5] Bikel, D. M., Miller, S., Schwartz, R., Weischedel, R. (1997). Nymble: a high-performance learning name-finder. In Proceedings of the fifth conference on Applied natural language processing. pp. 194-201. Association for Computational Linguistics.

[6] Borthwick, A., Sterling, J., Agichtein, E., Grishman, R. (1998). Exploiting diverse knowledge sources via maximum entropy in named entity recognition.

[7] McCallum, A., Li, W. (2003). Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In the Proceedings of HLT-NAACL. Vol. 4, pp. 188-191. Association for Computational Linguistics.

[8] Wu, Y.C., Fan, T.K., Lee, Y.S., Yen, S.J. (2006). Extracting Named Entities Using Support Vector Machines, Spring-Verlag, Berlin Heidelberg.

[9] Greenberg, J. (1966). The Languages of Africa. The Hague.

[10] Ouakrim, O. (1995). Fonética y fonología del Bereber. Survey at the University of Autònoma de Barcelona.

[11] Cunningham, H. (2002). Gate, a general architecture for text engineering. Computers and the Humanities. 36(2): 223–254.

[12] Maynard, D., Cunningham, H., Bontcheva, K., Catizone, R., Demetriou, G., Gaizauskas, R., Hamza, O., Hepple, M., Herring, P., Mitchell, B., Oakes, M., Peters, W., Setzer, A., Stevenson, M., Tablan, V., Ursu, C., Wilks, Y. (2000). A survey of uses of gate. Technical Report CS-00-06, Department of Computer Science, University of Sheffield.

[13] Talha, M., Boulaknadel, S., Aboutajdine, D. (2014). NERAM: Named Entity Recognition for Amazigh language. In: 21th International conference of TALN. pp. 517-524. Aix Marseille University, Marseille.

[14] Boulaknadel, S., Talha, M., Aboutajdine, D. (2014). Amazigh Named Entity Recognition Using a Rule Based Approach. In: 11th ACS/IEEE International Conference on Computer Systems and Applications. Doha, Qatar.

[15] Talha, M., Boulaknadel, S., Aboutajdine, D. (2015). L'apport d'une approche symbolique pour le repérage des entités nommées en langue amazighe. In: EGC. pp. 29-34. Luxembourg.

[16] Talha, M., Boulaknadel, S., Aboutajdine, D. (2018). Performance Evaluation of SVM-Based Amazigh Named Entity Recognition. In International Conference on Advanced Machine Learning Technologies and Applications. pp. 232-241. Springer, Cham.

[17] Cortes, C., Vapnik, V. (1995). Support-vector networks. In Machine Learning, pp. 273-297.

[18] Talha, M., Boulaknadel, S., Aboutajdine, D. (2015). Development of Amazigh Named Entity Recognition System Using Hybrid Method. Journal of Research in Computing Science. Vol. 90, pp. 151-161.

Text Mining : de l'Information Textuelle à la Connaissance

SELLAMY Khadija[1], FAKHRI Youssef[1], BOULAKNADEL Siham[2], MOUMEN Anis[3], HAFED Khalid[1]

[1] Laboratoire de Recherche en Informatique et Télécommunications LARIT, Faculté des sciences Kénitra, Université Ibn Tofail, Maroc

[2] Institut Royal de la Culture Amazighe, Allal El Fassi Avenue, Madinat Al Irfane, Rabat-Instituts, Maroc

[3] Laboratoire Génie des Systèmes, ENSA de Kénitra, Université Ibn Tofail, Kénitra, Maroc

{sellamyinfo,moumen.anis, hafed.khalid}@gmail.com

fakhri@uit.ac.ma

boulaknadel@ircam.ma

Résumé

Le Text Mining ou bien la fouille de texte est l'utilisation des techniques et méthodes automatisées pour la découverte des connaissances à partir des grandes bases de données textuelles (TDM) et de diverses origines : des articles de presse ou scientifiques, des entretiens, des questionnaires, des sites Web, pages réseaux sociaux, forums. Cet axe de recherche a changé la grandeur d'utilité des technologies de l'information textuelle de la reconquête à l'analyse et l'exploration automatique et rapide des textes, son objectif principal est de traiter le contenu des textes d'en extraire les informations parlantes. En effet, le Text Mining fournit des connaissances qui pourraient être utilisées pour la prise de décisions.

Dans ce travail, nous allons discuter les techniques et les méthodes de Text Mining, ses limitations et ses applications dans divers domaines ; nous allons voir les étapes à suivre pour préparer un corpus pour l'analyse textuelle. Ensuite, tout en s'intéressant aux outils d'analyse de données textuelles les plus connus et utilisés nous allons présenter un exemple d'étude en utilisant le logiciel IRAMUTEQ basé sur R et Python.

1. Introduction

Le Text mining « la fouille de textes » ou bien l'analyse des données textuelles est l'utilisation des techniques et méthodes automatisées pour la découverte des connaissances à partir des grandes bases de données textuelles (TDM) de diverses origines et de différents types (documents structurés, semi- structurés, et non structurés). Le TM est une variante du domaine de data mining qui a pour objectif d'extraire des modèles captivants, significatifs et des informations inconnues préalablement. Il utilise une variété d'outils d'analyse pour découvrir des régularités des relations dans les données.

Le TM est un champ multidisciplinaire basé sur plusieurs domaines, citons : la fouille de données (data mining), l'informatique, la linguistique, les techniques d'apprentissage automatique (machine Learning), et le web mining. La figure 1 montre l'interaction du text mining avec d'autres champs [1].

Figure1 : Interaction du text mining avec d'autres domaines

En statistique, le text mining renvois à l'exploitation de deux domaines : le data mining et le machine Learning, soit pour un objectif descriptif pour caractériser les propriétés générales des données cachées par le volume des données ou bien pour un objectif prédictif pour prédire des nouvelles informations à partir des informations présentes.

En linguistique, on cherche à développer des moyens de calcul pour répondre aux exigences scientifiques de la linguistique (Analyse lexicale, analyse syntaxique, NLP, …) tout en modélisant l'utilisation de la langue humaine par des modèles mathématiques. Le Web Mining est l'analyse comportementale des internautes. Il est composé de trois catégories, à savoir : (1) l'analyse du contenu des pages Web (Web Content Mining) est le processus d'extraction des connaissances à partir du contenu des pages Web ; (2) l'analyse des liens entre les pages Web (Web Structure Mining) et (3) l'analyse de l'usage des pages Web (Web Usage Mining) consistent à analyser le comportement des utilisateurs à travers l'analyse de leur interaction avec le site Web.

En général, Le processus du text mining suit les étapes suivantes [1] :

- La collecte de données à partir de différentes sources disponibles dans différents formats de fichiers.
- Le prétraitement et le nettoyage des données sont effectués pour détecter et éliminer les anomalies. Le processus de nettoyage veille à capturer l'essence réelle du texte disponible et il est exécuté pour supprimer les mots vides liés au processus (processus d'identification de la racine de certains mots) et pour indexer les données.
- Le traitement et le contrôle : ces opérations sont appliquées pour auditer et nettoyer davantage les données par un traitement automatique.

Les informations traitées dans les étapes ci-dessus sont utilisées pour extraire des informations précieuses et pertinentes pour la prise de décision et analyse de tendances futures.

2. Text Mining : approches et techniques

Différentes techniques du text mining sont appliquées pour analyser des modèles de textes, parmi ces techniques nous trouvons : Extraction de l'Information (EI), la Recherche d'Information (RI), la catégorisation, le Traitement du Langage Naturel (TLN) et la visualisation de l'information.

2.1. Extraction de l'information

L'Extraction de l'Information (EI) est une technique qui permet d'extraire des entités d'informations significatives et des attributs spécifiques à partir de grands volumes de textes. Le système d'IE a pour rôle non seulement d'extraire des entités et des informations approfondies mais aussi d'identifier les relations entre elles. Les informations requises sont stockées dans une base de données afin d'appliquer des techniques de DM pour avoir de nouvelles connaissances [2,3], comme indiqué dans la figure suivante.

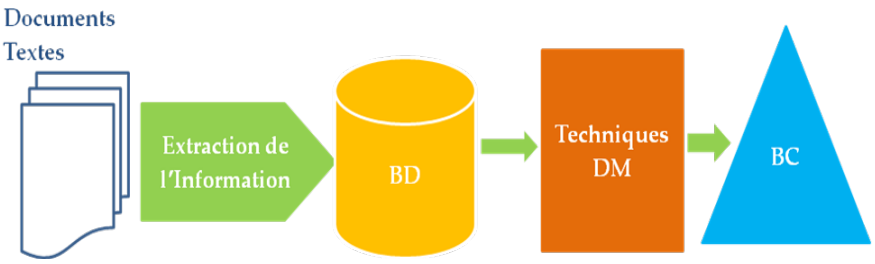

Figure 2 : Extraction de l'information

2.2. La recherche d'information

La Recherche d'Information (RI) est un processus d'extraction des informations particulières recherchées par l'utilisateur et qui se concentre sur la requête posée par ce dernier [4]. Les moteurs de recherche Google et Yahoo utilisent souvent le système de recherche d'informations pour extraire les documents pertinents, en utilisent des algorithmes basés sur des requêtes pour fournir aux utilisateurs des informations appropriées à leurs besoins.

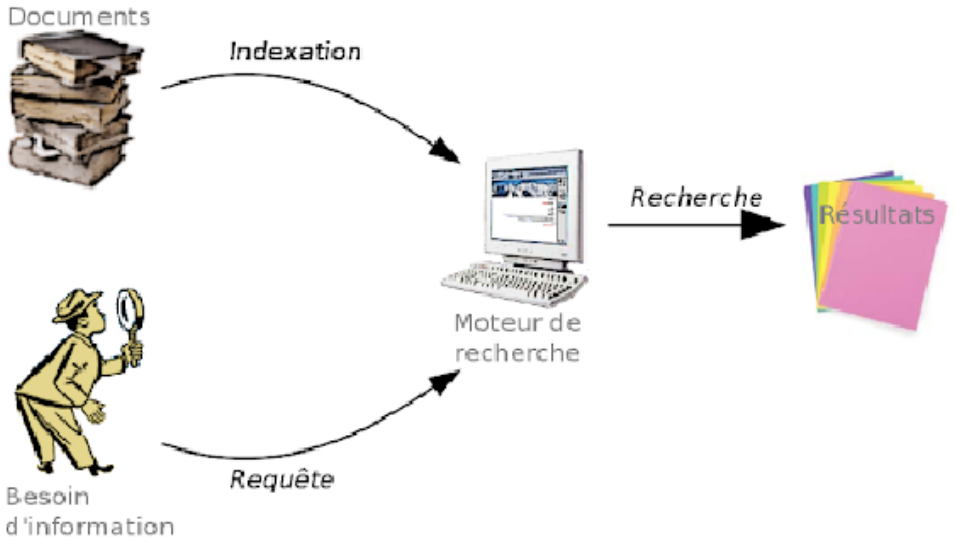

Figure 3 : Recherche d'information

2.3. La catégorisation

La catégorisation automatique des textes est une technique qui consiste à affecter des documents à des catégories prédéfinies en fonction de leur sujet et leur contenu. Elle identifie les thèmes principaux d'un document en le plaçant dans un ensemble prédéfini de sujets. Lors de la catégorisation d'un document, un programme informatique le traite souvent comme un "sac de mots", Il ne traite pas les informations de même comme le système d'extraction d'informations [4 ; 5].

2.4. Traitement du langage naturel

Traitement du langage naturel est un domaine au carrefour de l'informatique, de l'intelligence artificielle et de la linguistique qui vise à la compréhension du langage naturel en utilisant la machine. Le TLN est intéressé par la Génération de Langage Naturel (GLN) et la Compréhension du Langage Naturel (CLN). Le système GLN sert à des représentations linguistiques et un déclencheur syntaxique pour s'assurer que le texte généré est grammaticalement correct, afin de respecter les règles grammaticales et la structure du texte : les phrases, les paragraphes sont arrangés d'une manière cohérente. Le système CLN s'intéresse au sens du texte, il comprend au moins un des composants suivants : la tokenisation, l'analyse morphologique ou lexicale, analyse syntaxique et analyse sémantique [3].

2.5. Visualisation de l'information

La visualisation de l'information permet de présenter de grandes quantités des données textuelles dans une carte visuelle. Cette technique permet aussi à l'utilisateur d'analyser visuellement le contenu et elle fournit des différentes fonctionnalités de navigation et de recherche d'information. L'utilisateur peut interagir avec la carte du document en effectuant un zoom, une mise à l'échelle et en créant des sous-cartes [2 ; 6].

3. Text Mining : domaines d'application

Le text mining a une grande valeur et un plus pour divers domaines, vu qu'il permet d'explorer une grande quantité de données textuelles pour extraire des modèles intéressants :

- Business Intelligence ;
- Médias sociaux ;
- Domaine académique et la recherche Scientifique ;
- Application de sécurité ;
- Gestion des ressources humaines ;
- Amélioration de la recherche Web ;
- Bibliothèques numériques ;
- Sociétés pharmaceutiques et la recherche de santé.

4. Text Mining : outils

De nos jours, il existe une variété d'outils de text mining utilisés pour l'analyse de données textuelles et l'extraction de nouvelles informations. Dans cette partie, nous allons présenter le logiciel IRAMUTEQ et un exemple d'étude.

4.1. Présentation du logiciel

IRAMUTEQ est un logiciel d'analyse des données textuelles. Il s'appuie sur le logiciel R et sur le langage python. Iramuteq propose un ensemble de traitements et d'outils pour l'aide à la description et à l'analyse de corpus texte. Il permet d'analyser des corpus volumineux au format texte brut (.txt), qui peut représenter un entretien, un article ou tout autre type

de documents texte. Ce logiciel donne la possibilité de combiner des analyses de contenu et des analyses statistiques.

Dès l'ouverture du corpus, IRAMUTEQ nous propose des consignes de préparation de ce dernier. Il affiche une boîte de dialogue d'indexation pour définir les caractéristiques générales, ainsi que les différentes options d'indexation, plus l'onglet "nettoyage" pour (Passer le corpus en minuscule, remplacer les apostrophes par des espaces, remplacer des tirets par des espaces, conserver la ponctuation, éliminer d'espace entre deux formes), ensuite la lemmatisation et l'identification des clés d'analyses ou bien des formes actives.

Figure 4 : Interface du logiciel IRAMUTEQ

IRAMUTEQ propose plusieurs outils d'analyse :

- Analyse statistique textuelle ;
- Spécificité et analyse factorielle des correspondances (AFC) ;
- Méthode de classification Reinert ;
- Analyse de similitude ;
- Nuage de mots.

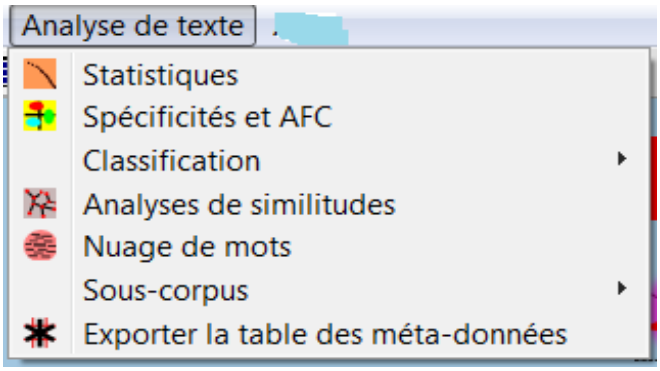

Figure 5 : Onglet d'analyse de texte

4.2. Présentation d'un exemple traité par IRAMUTEQ

Dans cette partie, nous allons présenter un exemple d'analyse de données textuelles pour illustrer certaines fonctionnalités du logiciel IRAMUTEQ, l'exemple est effectué sur un corpus d'un ensemble de résumés (Abstracts) des articles scientifiques. Le corpus est constitué de 35 résumés et de 5416 mots.

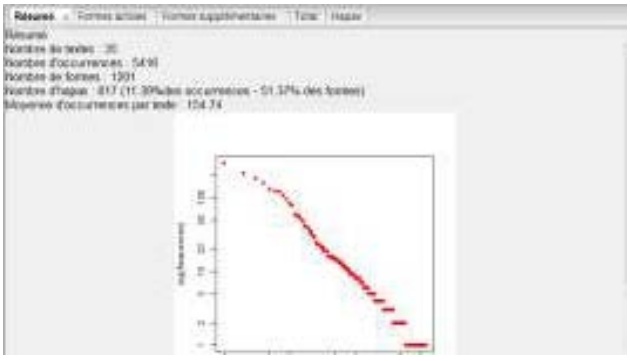

Figure 6 : Caractéristiques générales du corpus

4.2.1. Les mots clés des résumés : nuage des mots

Dans la technique du nuage de mots, la fréquence d'occurrence des mots détermine leur taille. D'après la figure 7, nous pouvons savoir quels sont les mots-clés les plus fréquents dans les résumés collectés, nous remarquons que "data mining" est le plus mentionné, en deuxième le "Web mining" et en troisième place nous trouvons le 'text mining". Ensuite, les mots "techniques", "information", "analysis", "paper" respectivement. Nous

remarquons aussi que le terme "mining" a la plus grande taille au fait qu'il est toujours attaché aux termes "data", "web" et "text" presque cité dans tous les résumés. L'apparition de ces mots nous indique que notre corpus est constitué des résumés des articles qui se traitent les 3 domaines suivants : data mining, text mining, le web mining et leurs techniques.

Dans la figure 8, nous remarquons un changement au niveau des termes, par exemple au lieu du "mining" nous trouvons "mine", s'est dû au fait de la lemmatisation. Cette méthode sert à identifier la racine d'un mot en éliminant divers suffixes, les différentes variantes d'un mot sont regroupées pour réduire le nombre de mots et d'économiser du temps [7]. Par exemple, les mots construire, construction, construit peuvent tous être associés à un seul mot représenté par la racine des variantes.

Figure 7 : Nuage de mots sans Lemmatisation

Figure 8 : Nuage de mots avec Lemmatisation

4.2.2. Les liens entre les mots : analyse des similitudes

Nous avons effectué l'analyse des similitudes sur l'ensemble des résumés collectés pour savoir quels sont les mots liés entre eux et quelle est la fréquence de ces liens. IRAMUTEQ nous propose différents modes de représentations graphiques des liens entre les mots et aussi la fréquence de ces liens.

Les Figures 9 et 10 montrent les relations entre tous les mots des résumés, nous trouvons par exemple le mot "mining" est au cœur du graphique et liés aux autres mots "data", "web", "text", "technique". L'épaisseur des arêtes nous indique les degrés de relation entre les termes. Dans la figure 11, nous constatons l'affichage des indices sur les arêtes, ce qui indiquent la fréquence des liens entre les mots liés. Ceci est attribué au fait que l'option "Indice sur les arêtes" est activé au niveau du paramétrage de l'analyse.

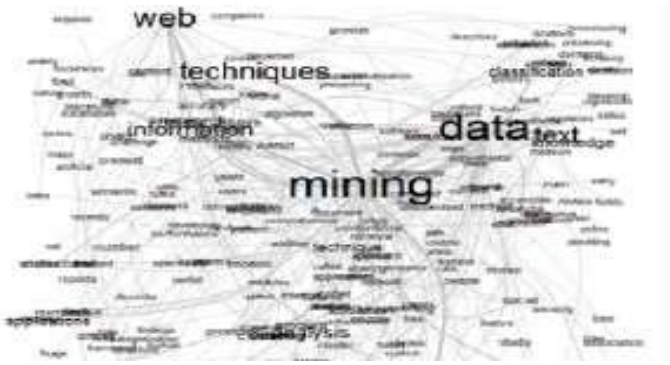

Figure 9 : Analyse des similitudes présentation "reingold"

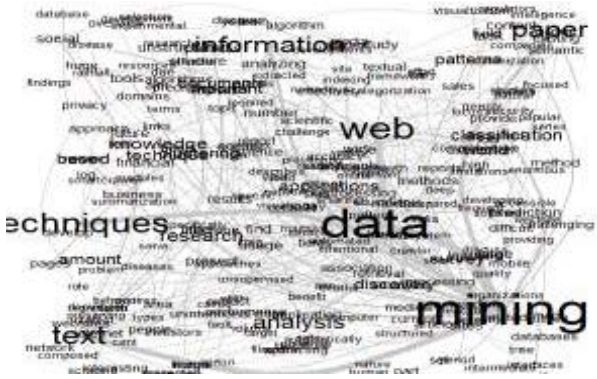

Figure 10 : Analyse des similitudes présentation "random "

5. Conclusion

Dans cet article, nous avons présenté le text mining, ses techniques et approches, ensuite, les étapes de base pour le processus de TM avec un aperçu sur certains domaines d'application. Nous avons, également, présenté un outil d'analyse de données textuelles IRAMUTEQ avec un exemple de corpus pour illustrer certaines fonctionnalités de ce logiciel.

Références

[1] Talib, R., Kashif, M., Ayesha, S., Fatima, F. (2016). Text Mining: Techniques, Applications and Issues, International Journal of Advanced Computer Science and Applications. Vol. 7, N° 11.

[2]. Thilagavathi, K, Shanmuga, V., PRIYA, M., Phil, M. (2014). A survey on Text Mining techniques, International Journal of Research in Computer Applications and Robotics. Vol.2, N° 10.

[3] Inzalkar, S. M., Sharma, J. (2017). A Survey on Text Mining- techniques and application, International Journal of Science and Research. Vol. 6.

[4] Janani, R., Vijayarani, S. (2016). Text Mining Research: A Survey, International Journal of Innovative Research in Computer and Communication Engineering. Vol. 4, N° 4.

[5] Gupta, V., Lehal, G. S. (2009). A Survey of Text Mining Techniques and Applications », Journal of Emerging Technologies in Web Intelligence. Vol. 1, N° 1.

[6] Patel, R., Sharma, G. (2014). A survey on Text mining techniques, International Journal of engineering Computer Science. Vol.2, N° 5.

[7] Vijayarani, G., Kannan, S. (2014). Preprocessing Techniques for Text Mining - An Overview, International Journal of Computer Science & Communication Networks. Vol. 5, N°1, pp. 7-16.

[8] Salloum, S., Al-Emran, M., Abdel Monem, A., Shaalan, K. (2018). Using Text Mining Techniques for Extracting Information from Research Articles, Studies in Computational Intelligence. pp. 373.

[9] Sukanya, M., Biruntha, S. (2012). Techniques on text mining, Advanced Communication Control and Computing Technologies (ICACCCT), IEEE International Conference. pp. 269-271.

[10] Sagayam, R. (2012). A survey of text mining: Retrieval, extraction and indexing techniques, International Journal of Computational Engineering Research. Vol. 2, N° 5.

[11] Boughzala, Y. (2014). Sphinx Quali : un nouvel outil d'analyses textuelles et sémantiques, 12es Journées internationales d'Analyse statistique des Données Textuelles.

[12] Niharika, S., Sneha Latha, V., Lavanya, D.R. (2012). A survey on text categorization, International Journal of Computer Trends and Technology. Vol. 3, N°1.

[13] Nasa, D. (2012). Text Mining Techniques - A Survey, International Journal of Advanced Research in Computer Science and Software Engineering. Vol. 2, N° 4, pp. 551-540.

[14] Saranya, A. S., Geetharani, S. (2016). A Study on Web Mining Tools and Techniques, International Journal of Innovative Research in Computer and Communication Engineering. Vol. 4, N° 4.

[15] Zhong, N., Li, Y., Wu, S.T. (2012). Effective pattern discovery for text mining, IEEE transactions on knowledge and data engineering. Vol. 24.

[16] Documentation_iramuteq_21 21_12_2013 http://www.iramuteq.org/news .pdf.

Automatic Recognition of Arabic Handwritten Words

LAMSAF Asmae[1], AIT KERROUM Mounir[1], BOULAKNADEL Siham[2], SAMIR Chafik [3]

[1] LaRIT, Faculty of Sciences- Ibn Tofail University, Kenitra, Morocco

[2] Royal Institute of Amazigh Culture -Rabat, Morocco

[3] University of Clermont Auvergne, France
{asmaelamsaf, maitkerroum} @gmail.com
boulaknadel@ircam.ma
chafik.samir@udamail.fr

Abstract

In this paper, we present a system of automatic recognition of Arabic handwritten words. The proposed system is based on global approaches, which consist in recognizing all the word without segmenting into the characters in order to recognize them separately. The classifier used in this work is based on the principle of the K-nearest-neighbors approach, which based to compute the distance between the feature vectors of the test images and the images of the database. After the pretreatment of the word image, we propose to use the result of the zoning technique to build the feature vectors. These vectors are considered as input of the classification step of the test images. The recognition rate of our system achieves 89% on the IFN / ENIT database.

1. Introduction

The understanding of writing by a computer is still far from being fully satisfactory, the reason is related to the fact that the study of the recognition of the writing is a very vast field as by its applications as by its techniques. Therefore, it is necessary to propose adapted treatment a method, among which is the automatic recognition of handwriting, which allows transforming handwriting into its symbolic representation understandable

by a machine and easily reproducible by a computer system. The automatic recognition of the writing is attached to the recognition of characters manuscripts, the recognition of words and text. We choose the Arabic language as the language of practice of our method.

The recognition of Arabic handwritten words is an important step to realize a system of recognition of Arabic handwritten texts. The main role of word recognition is to transform of a handwritten word to its symbolic representation using a database of words.

The general problem of the recognition of handwritten words is the infinity of representation of writing because each person has his own style of writing. It is necessary to distinguish also the two approaches of the word recognition, which raises the global approaches and the analytical approaches. The analytical approach that addresses word recognition is to try to recognize each of these letters that compose them. The global approach consists of a global recognition of the word that considers the word as a whole without trying to locate each of the letters that compose it. In our work, we used the global approach without segmenting the images of words into characters.

There are several classifiers used for recognition of Arabic manuscript words or characters in the literature such as Hidden Markov Models (HMM) methods [1 ; 2 ; 3 ; 4 ; 5 ; 6 ; 7 ; 13], K nearest-neighbors [8 ; 9], neural networks [11 ; 12 ; 15], machine vector support (SVM) [10] and others [14]. Despite the efforts that have been made to achieve Arabic handwriting words recognition systems, recognition remains important research issues with regard to the operations used in this field that pose problems of execution time and some result feasible.

In this paper, we propose a method of recognition of Arabic handwritten words based on the principle of k-nearest-neighbors which consists in transforming the images of handwritten words into its symbolic representations. Our proposed system follows several steps like all recognition systems: preprocessing (binarization), feature extraction, classification using the IFN / ENIT database.

2. Arabic language

Arabic is written by more than 250 million people. Although the spoken Arabic is somewhat different from one country to another, the writing system is a standard version used by all the Arab people for their communication. Unlike other languages such as the Latin, Chinese, and Japanese scripts that are widely discussed, the recognition of Arabic handwritten text remains a challenge. The Arabic script has the following characteristics:

- By nature, the Arabic script is cursive,
- The Arabic text is written from right to left,
- More than half of the Arabic letters are composed of a main body and secondary components
- The type and position of the secondary component are very important features
- Most letters of Arabic words are linked together, depending on their position (Start, Middle, End)
- There are a small number of letters that have the same shape regardless of the position.
- Three types of words in Arabic (words that contain: separate letters, connected components, linked letters).
- On the other hand, the breadth of Arabic letters differs from one letter to another; in handwritten scripts. In addition, usually there are differences in forms of letters from one person to another
- Vowels are not letters, but diacritics associated with the letters to which they apply.

sound	name	name	end	middle	start	isolated
ā	alif	ألف	ـا	ـا*	ا*	ا
b	bā’	باء	ـب	ـبـ	بـ	ب
t	tā’	تاء	ـت	ـتـ	تـ	ت
th	thā’	ثاء	ـث	ـثـ	ثـ	ث
dj	djīm	جيم	ـج	ـجـ	جـ	ج
H	Hā’	حاء	ـح	ـحـ	حـ	ح
kh	khā’	خاء	ـخ	ـخـ	خـ	خ
d	dāl	دال	ـد	ـد*	د*	د
dh	dhāl	ذال	ـذ	ـذ*	ذ*	ذ
r	rā’	راء	ـر	ـر*	ر*	ر
z	zāy	زاي	ـز	ـز*	ز*	ز
s	sīn	سين	ـس	ـسـ	سـ	س
sh	shīn	شين	ـش	ـشـ	شـ	ش
S	Sād	صاد	ـص	ـصـ	صـ	ص
D	Dād	ضاد	ـض	ـضـ	ضـ	ض
T	Tā’	طاء	ـط	ـطـ	طـ	ط
Z	Zā’	ظاء	ـظ	ـظـ	ظـ	ظ
c	ᶜayn	عين	ـع	ـعـ	عـ	ع
gh	ghayn	غين	ـغ	ـغـ	غـ	غ
f	fā’	فاء	ـف	ـفـ	فـ	ف
q	qāf	قاف	ـق	ـقـ	قـ	ق
k	kāf	كاف	ـك	ـكـ	كـ	ك
l	lām	لام	ـل	ـلـ	لـ	ل
m	mīm	ميم	ـم	ـمـ	مـ	م
n	nūn	نون	ـن	ـنـ	نـ	ن
h	hā’	هاء	ـه	ـهـ	هـ	ه
w, ū	wāw	واو	ـو	ـو*	و*	و
y, ī	yā’	ياء	ـي	ـيـ	يـ	ي

Figure 1: Arabic alphabet

3. Architecture of the proposed system

Our system based on the global approach of recognizing the word without segmenting into characters. The input data of our recognition system are images that contain handwritten words written by several people in a different way. We applied preprocessing (binarization, normalization) to prepare the images for the next phase, after the preprocessing step we will

extract the characteristics of each image as vectors to classify these vectors using the database of handwritten words IFN / ENIT.

3.1. Preprocessing

Preprocessing consists of preparing the data from the sensor for the next phase. It's basically about reducing the noise of the data and trying to keep only the meaningful information of the form represented. The operations used in our work: Binarization, normalization.

3.1.1. Binarization

For the binarization, we opted for a global thresholding which consists in taking an adjustable threshold, but identical for the whole image. Each pixel of the image is compared with this threshold and takes the value white (= 1) or black (= 0) according to whether it is higher or lower.

Figure 2: (a) Input image, (b) Binarized image (output image)

In Figure 2, we present the input data of system of binarization is image of handwritten word and output image is binarized image of word with the writing in white and the background in black.

3.1.2. Normalization

After the previous step, it is necessary to standardize the size of images. The normalization is an operation that can bring the words images to standard forms concerning the size of the image. All resulting images of this step are the same size.

3.2. Feature extraction

One of the fundamental problems of pattern recognition is to determine what characteristics to use to get the correct result from the classification. In the context of this work, we use the statistical characteristics derived from the pixel distribution as zoning [14].

The steps of the zoning method can be summarized in three steps:

- Compute the total number of white pixels in the image (the pixels of the writing).
- Divide the image into a x a zones of the same size.
- For each zone, we compute the average value of white pixels as:

$$C_j = R_j / N$$

Where C_j: the result value in the zone j.

R_j: number of white pixels in the zone j.

N: total number of white pixels of the word.

The result obtained will be used as a feature vector to represent the input image in the classification step.

3.3. Classification

After the feature extraction phase that gives the feature vectors of the word images, these characteristics should be classified to recognize the set of words. The role of a classifier is determined among a finite set of classes to which a given object belongs. In this work, we use the principle of KNN which consists in locally evaluating the similarity between the words, or more precisely evaluating the distance between the feature vectors of the test words and the feature vectors of the words of the database.

The classification can be divides into two steps: the training which classifies the words of the learning base according to their feature vectors to initialize the model database and then the test or decision which consists in assigning a class for each new observation. The classification process consists of determining which class an input image belongs to.

3.3.1. The training

This phase consists of initializing or creating the base of the models by saving the characteristic vectors of the different words of the database. The features of the previous step are used to identify an image of a word. The steps of this method are:

- Apply the zoning algorithm for each image of the database: divide the image into blocks of a x a (a = 5, 10, ...) and sum the pixels of each block and divide on a x a to obtain a vector of features that contains the sum of pixels of all blocks in the image.
- Create a matrix that groups the feature vectors of the images in the database. The first column of the matrix contains the first image feature vector, the second column contains the second image feature vector, and so on until all the images of the database.
- At the end this phase, the final matrix called model base that stores all the feature vectors of the words of the database.

3.3.2. The test phase

It consists in using the characteristics extracted to assign a class based on the data of the base of the models. Note that the steps applied to the database images in the training phase are applied to the test images to extract the feature vectors from the test images.

This phase consists in testing the distance between the feature vector of the test input images (image contains the test word) and the model base of the preceding step, to do this, we compare the feature vector of the input word with all the columns of the model base, then we determine the sum of each comparison vector, and we choose the four minimums of the resulting vector, the minimum number of columns of the vectors is the number of the solution images of the recognition.

At the end of the algorithm, we get the four columns that represent the recognition system output images. These images are characterized by several characteristics given in the database that can be used to extract the symbolic representation of the Arabic word (city code, name...).

4. Results

As part of this work, we realized an adaptive method for automatic recognition of Arabic handwritten words which aims to transform the images of Arabic handwritten words into its symbolic representation using the IFN / ENIT database [13]. The database is composed of 5 subsets (in version v2.0p1e) with 32492 images of names of Tunisian cities / villages collected from more than 1000 writers. The number of test image is 80 images and the number of used image of the database is 569.

To evaluate the performance of our Arabic handwriting recognition system, we computed the recognition and results are summarized in tables 1, 2 and 3.

a=5	
Number of minimums	Recognition rate
1	35%
2	45%
3	45%
4	55%

Table 1: Recognition rate for the values of a =5

a=10	
Number of minimums	Recognition rate
1	55%
2	65%
3	70%
4	89%

Table 2: Recognition rate for the values of a=10

a=20	
Number of minimums	Recognition rate
1	45%
2	62%
3	68%
4	80%

Table 3: Recognition rate for the values of a=20

In tables 1, 2 and 3, we present the comparison of recognition rate with several choices of number of the output images, we find that when increasing the number of minimum (the resulting image number) the recognition rate increases with the number of blocks a = 5, a = 10 and a = 20.

After comparing recognition rates for several values of a, we selected the value of a better results obtained which is the result of the data of a = 10 with 4 minimums.

This method has the advantage of being simple to implement, it gives acceptable results. The disadvantage of this method is the running time is long to compute the base model matrix.

5. Conclusion

We have presented a system for automatic recognition of Arabic handwriting words, we have detailed the steps that we used to realize our system as pretreatment, feature extraction and classification. We use pixels' distributions to build our features' vectors and a KNN classifier. The proposed method has been tested and compared on the IFN / ENIT international database, and the results obtained are encouraging with 89% recognition rate.

Références

[1] Alkhoury. I. (2010). Arabic handwritten word recognition based on Bernoulli mixture HMM. Master Thesis, University of Valencia, Spain.

[2] Kessentini, Y., Paquet, T., Burger, T. (2010). Comparaison des méthodes probabilistes et évidentielles de fusion de classifieurs pour la reconnaissance de mots manuscrits, Proceedings du Colloque International Francophone sur l'Écrit et le Document (CIFED), Sousse-Tunisie.

[3] Alkhateeb, J.H., Pauplin, O., Ren, J., Jiang, J. (2011). Performance of hidden Markov models and dynamic Bayesian network classifiers on handwritten Arabic word recognition, Knowledge-Based System, Elsevier. Vol. 24, pp. 680-688.

[4] Alkhateeb, J.H. (2011). Word-based handwritten Arabic scripts recognition using dynamic Bayesian network, Proceedings of the 5th International Conference on Information Technology (ICIT), Amman-Jordan.

[5] Burrow, P. (2004). Arabic handwriting recognition, Master Thesis, School of Informatics, University of Edinburgh, England.

[6] Farah, N., Souici, L., Sellami, M. (2006). Classifiers combination and syntax analysis for Arabic literal amount recognition, Engineering Applications of Artificial Intelligence, Elsevier. Vol. 19, pp. 29-39.

[7] Khalifa, M., Ru, Y.B. (2011). A novel word based Arabic handwriting recognition system using SVM classifiers, Advanced Research on Electronic Commerce, Web Application, and Communication (ECWAC), Springer. Vol. 143, pp. 163-171.

[8] Nouar, F., Aissaoui, M.E., Seridi, H. (2008). Approche globale pour la reconnaissance de mots arabes manuscrits par combinaison parallèle de classifieurs, Proceedings des Journées des Jeunes Chercheurs en Informatique (JCI), Guelma-Algeria.

[9] Ebrahinpour, R., Amini, M., Sharifizadehi, F. (2011). Farsi handwritten recognition using combining neural networks based on stacked generalization, International Journal of Electrical Engineering and Informatics. 3(2):146-160.

[10] Khémiri, A., KacemEchi, A., Belaid, A., Elloumi, M. (2016). A System for off-line Arabic Handwritten Word Recognition based on Bayesian Approach, 15th International Conference on Frontiers in Handwriting Recognition.

[11] Abuzaraida, M. A., Zeki, M., Zeki, A. M. (2017). Online Recognition of Arabic handwritten words system based on Alignments matching Algorithm, proceedings of the International conference on computing, Mathematics and statistics, Springer Nature Singapore.

[12] Ouchtati, S., Redjimi, M., Bedda, M. (2014). Recognition of the Arabic Handwritten Words of the Algerian Departments, International Journal of Computer Theory and Engineering. Vol. 6, No. 2.

[13] El Abed, H., Märgner, V. (2007). The IFN/ENIT-database - a tool to develop Arabic handwriting recognition systems, in 9th International Symposium on Signal Processing and Its Applications. pp. 1-4.

[14] Rajabi, M., Nematbakhsh, M., Monadjemi, N. (2012). A New Decision Tree for Recognition of Persian Handwritten Characters, International Journal of Computer Applications. 44(6):52-58.

[15] Lawgali, A. (2015). A Survey on Arabic Character Recognition, International Journal of Signal Processing, Image Processing and Pattern Recognition. 8(2):401-426.

Boulid, Y., El Youssfi, E. M., Souhar, A. (2016). Reconnaissance de l'écriture manuscrite arabe en mode hors ligne.

تتمحور الأعمال المنشورة في هذا الكتاب حول مسألة أساسية، تتمثل في معالجة اللغات من خلال العلوم الرقمية، حيث تتناول مجالين رئيسيين : التعلم بوساطة التكنولوجيا، وتطوير الموارد والأدوات اللغوية. وهذه الأعمال مساهماتٌ لباحثين ومهنيين، وطنيين ودوليين، يعملون في مجال العلوم الرقمية المطبقة على اللغات الطبيعية.

* * *

ⵙⵓⵛⵖ++ⵓ ⵖⴰⵎⵙⵓ ⵄⴰ +ⵙⵍⵖⵓⵙⵍⵉⵙⵉ ⵙⵓⵄⵍⵄⵎⵏⵉ ⵅ ⵣⵙⵆ+ ⵏ +ⵛⵓⵎⵄⵣ+ +ⵄⴰⴰⵓⵎⵏⵉ+ ⵏ ⵖⵓⵛⵔⵎ ⵏ +ⵖ+ⵎⵄⵣⵙⵉ ⵓ +ⵛⵄⵓⵓⵓⵏⵉⵙⵉ +ⵙⵛⵖⵋⵋⵖⵏⵉⵙⵉ.

ⵄⵣⵄ ⵏ +ⵍⵍⵖⵓⵙⵍⵉⵙⵉ ⵓⵙⵢⵙⴰⵏ+ ⅹ ⵓⵙⵉ ⵏ ⵙⵅⵓⵄⵏ : ⵄⵎⵛⵛⵖⴰ ⵓ +ⵙⵔⵏⵖⵎⵖⵉⵙ+ ⴰ ⵖⵓⵜⵖⵢⵎⵖ ⵏ ⵙⵓⵖⵅⵄⵛ ⴰ ⵙⵛⵄⵓⵓⵏⵉ.

+ⵙⵍⵖⵓⵙⵍⵉⵙⵉ ⵄⴰ ⵓⵖⵛⵏ+ ⵄⴰⵓⵄⵍ ⵏ ⵙⵛⵖⵅⵄⵣⵉ ⴰ ⵙⵛⵓⵅⵖ+ⵏ ⵙⵏⵛⵖⵓⵏ ⴰ ⵙⵅⵓⵄⵢⵎⵄⵏ ⵅ ⵣⵙⵅⵓ ⵏ +ⵛⵄⵓⵓⵓⵏⵉⵙⵉ +ⵙⵛⵖⵋⵋⵖⵏⵉⵙⵉ +ⵖⵓⵏⵉⵙⵓⵙⵉ ⅹⵃ +ⵖ+ⵎⵄⵣⵙⵉ +ⵙⵅⵄⵛⵓⵏⵉⵙⵉ.

* * *

Les travaux présentés dans ce recueil portant sur la question essentielle du traitement des langues par les sciences du numérique. Ils abordent deux thèmes majeurs, à savoir l'apprentissage médiatisé par la technologie, et l'élaboration de ressources et outils linguistiques.

Ces travaux consistent en contributions des scientifiques, chercheurs et professionnels nationaux et internationaux, qui œuvrent dans le domaine des sciences du numérique appliquées aux langues naturelles.

34 DH